統計数理(2025)
第 73 巻 第 2 号 287–316

特集「生存時間解析と信頼性解析」
[総合報告]

極値統計論に基づくモデリング

吉田 拓真[1]・北野 利一[2]

(受付 2024 年 11 月 29 日;改訂 2025 年 5 月 13 日;採択 5 月 26 日)

要 旨

豪雨や地震などに起因する自然災害,ファイナンスにおける金融リスク,そして製品寿命など様々な現象に関するデータを扱う応用分野において,データ全体の中で極めて大きい,または小さい値の発生確率の見積りは信頼性評価やリスク管理の観点から重要な課題である.この課題に対する統計学的なテーマはデータの最大値や最小値,あるいはそれらに近い分位点の予測となる.そのためには,裾の挙動にのみ焦点を当てた確率モデルを構築することが不可欠である.極値統計学はそのための方法論を示すものであり,その稀な事象が生起する分布の裾のモデリングに数理的に重要な枠組みを与える.本稿では極値統計学の基本的な考え方,統計理論,モデリングについて総説する.特に,極値統計手法で実データ分析を行ったときに,その解析結果を現象の頻度特性の把握へとフィードバックする際に重要な役割を担う概念である再現期間と再現レベルは少し掘り下げた解説を行う.また,気象データや株価データなど極値統計学が活躍するデータはクラスターデータの形式で得られることが多い.そのため,クラスターデータに対する極値統計モデリングの方法論についても議論する.本稿では統計ソフト R のパッケージを紹介しながら降雨量データへ極値統計モデリングの適用例を示す.

キーワード:一般化極値分布,一般化パレート分布,極値統計学,クラスターデータ,空間データ,再現期間.

1. はじめに

実用科学の様々な分野で得られるデータには時に極端な値が含まれる.その極端な値が起きる原因の特定は難しいことが多く,外れ値として扱われ,解析から取り除かれることも多い.しかし,自然災害対策や製品開発現場では,そういった極端な値が起きる確率を正しく見積もることが重要となることもある.極端なデータとはデータ全体を見ると分布の裾に位置するようなものを指す.実際に,データ全体は正規分布で表現できても裾の部分にブレがあるような例は多い.そもそもデータのボリュームゾーンと裾を同一の分布(モデル)で表現することは難しい.よって,極端な値に相当する分布の裾の予測にはそれに特化したモデリングが重要となる.極値統計学はデータ全体の分布の裾の挙動にのみ焦点を当て,その部分を取り出して新たな確率分布を再構築するための手法である.

極値統計学を用いた裾の予測モデルとして大きく分けて二つの確率分布が用いられる.一つは一般化極値分布,もう一つは一般化パレート分布である.一般化極値分布はいわゆるデータ

[1] 鹿児島大学 大学院理工学研究科:〒 890–0065 鹿児島県鹿児島市郡元 1–21–35; yoshida@sci.kagoshima-u.ac.jp

[2] 名古屋工業大学 大学院工学研究科:〒 466–8555 愛知県名古屋市昭和区御器所町; kitano@nitech.ac.jp

の最大値の確率分布を記述し，一般化パレート分布はデータ全体のうちの閾値超過データの確率分布を表現する．2 つのモデルで共通する特徴は分布の裾の挙動は大きく三つのタイプに分類されることである．一つは正規分布や指数分布のように指数関数の速度で裾が減衰する指数テール，もう一つはベータ分布や一様分布のように有限のエンドポイントが存在するようなショートテール，最後の一つは t 分布やパレート分布のように極端な値が発生しやすいヘビーテールである．興味ある現象のリスク評価を行う際に，分布の裾が上の三つのどのタイプに属するかを知ることは重要であろう．多くの統計モデリングはデータに対して特定の分布モデルを想定するところからスタートするが，それはこの裾のタイプを一つに絞っていることになる．極端な値の発生メカニズムをモデル化するにあたり，どの裾タイプを持つか？から調べることは極値統計論を適用する応用上の観点からは重要となる．一般化極値分布，一般化パレート分布はこれらの裾のタイプを一つのパラメータで統一的に表現しているので，このパラメータを推定することで裾タイプの判定から予測モデル構築までを一連に行うことができる．

本稿では，分布の裾挙動をモデリングするための極値統計学の基礎理論，予測モデリングの構築，結果の解釈の仕方について先行研究のサーベイ，実データ適用を交えて総説する．特に，結果の解釈として再現レベル，再現期間と呼ばれる重要な概念があるが，これらの概念は極値統計モデルを現象の頻度特性の把握にフィードバックする際に重要な役割を担うため，少し掘り下げた解説を行う．第 2 章ではデータの最大値が持つ分布として一般化極値分布を用いたモデリングを解説する．閾値超過部分のモデリングとして有用な一般化パレート分布は第 3 章で議論する．第 4 章では一般化極値分布，一般化パレート分布の基礎モデルの発展についてサーベイする．さらに，気象データ，金融データなど極値統計学分野が活躍する分野はデータはクラスターデータの形で与えられることが多い．第 5 章ではクラスターデータに対する極値統計モデリングについても議論する．第 6 章でまとめと，本稿で述べることが出来なかった多変量極値分布について簡単に触れる．

2. 一般化極値分布

2.1 一般化極値分布（Generalized Extreme Value distribution, GEV）

確率変数 $X_1, \ldots, X_n$ は独立同一に分布関数 F に従って発生されたとする．つまり，$F(x) = P(X_i \leq x)$ である．このとき，最大値 $\max_{1 \leq i \leq n} X_i$ の分布は

$$\begin{aligned} P\left(\max_i X_i \leq x\right) &= P(X_1 \leq x, \ldots, X_n \leq x) \\ &= \prod_{i=1}^{n} P(X_i \leq x) \\ &= F^n(x) \end{aligned}$$

となり，一般に $F(x) < 1$ である x に対して $F^n(x) \to 0 (n \to \infty)$ と 0 に退化してしまう．一方で，最大値に対して位置と尺度を n に応じてうまく調整した場合は有限の確率分布を持つことが知られている．Fisher and Tippet (1928)，Gnedenko (1943) は適当な分布 F に対して，数列 $\mu_n = \mu_n(F)$, $\sigma_n = \sigma_n(F)$, 分布関数 G が存在し，

$$P\left(\frac{\max_i X_i - \mu_n}{\sigma_n} \leq x\right) = F^n(\sigma_n x + \mu_n) \to G(x)$$

が成立することを示し，さらに G が F の条件によって三つのタイプに分類されることを示した．Jenkinson (1955) はその三タイプの確率分布を統一的に表現した．すなわち，G は分布関数 F に依存するパラメータ $\gamma = \gamma(F) \in \mathbb{R}$ を用いて，

$$G(x) = G_\gamma(x) = \exp[-(1+\gamma x)_+^{-1/\gamma}] \tag{2.1}$$

と表される．ただし，$(x)_+ = \max\{x, 0\}$ である．また，$\gamma = 0$ のとき，$G_0(x) = \lim_{\gamma\to 0} G_\gamma(x) = \exp[-e^{-x}]$ となる．実際に，(2.1)は γ の符号に応じて

$$G_\gamma(x) = \begin{cases} \exp[-(1+\gamma x)^{-1/\gamma}], & \gamma > 0 \quad \text{(F)} \\ \exp[-(1+\gamma x)_+^{-1/\gamma}], & \gamma < 0 \quad \text{(W)} \\ \exp[-e^{-x}], & \gamma = 0 \quad \text{(G)} \end{cases}$$

と三つのタイプに分類される．タイプ(F)はフレッシェタイプ(Fréchet type)，(W)は逆ワイブルタイプ(reversed Weibull type)，(G)はガンベルタイプ(Gumbel type)と呼ばれる．一般に，$\gamma > 0$ の(F)はヘビーテール，$\gamma < 0$ に対応する(W)はショートテール，$\gamma = 0$ の(G)は指数テールである．特に，タイプ(W)は $1 + \gamma x > 0$ を満たす必要があるため，有限エンドポイント $x < -1/\gamma$ が存在する．この統一表現されたモデル(2.1)は一般化極値分布(Generalized Extreme Value distribution, GEV)と呼ばれている．実データ分析では F が未知のため，当然 γ とその符号も未知である．データ分析では，基本的に極値分布のタイプを事前に指定することができないため，三タイプの極値分布の統一表現はデータ分析を行う上で大変有用であろう．一般に，分布関数 F がパラメータ γ の極値分布 G_γ を持つとき，$F \in \mathcal{D}(G_\gamma)$ と表される．

ここで，(2.1)を書き直すと，

$$P\left(\max_i X_i \le x\right) = F^n(x) \approx G_\gamma\left(\frac{x-\mu_n}{\sigma_n}\right)$$

と近似できることから，(μ_n, σ_n) もパラメータ (μ, σ) と見ると

$$P\left(\max_i X_i \le x\right) \sim G_\gamma\left(\frac{x-\mu}{\sigma}\right) = \exp\left[-\left(1+\gamma\frac{x-\mu}{\sigma}\right)_+^{-1/\gamma}\right]$$

で近似できることになる．右辺の分布を GEV(μ, σ, γ) と書き，μ を位置パラメータ，σ を尺度パラメータ，γ を形状パラメータと呼ぶ．GEV はデータの最大値の確率分布をモデリングすることで，元の分布 F の裾を表現するモデルとみなせる．

次に一般化極値分布を理解する上で，特に重要な性質を挙げる．自然数 h について，hn 個の確率変数 $X_1, \ldots, X_{hn}$ は独立同一に分布関数 F に従うとする．また，$F \in \mathcal{D}(G_\gamma)$ とする．このとき，n が十分に大きいとすると，ある (μ_{hn}, σ_{hn}) が存在して

$$P\left(\max_{1\le i\le hn} X_i \le x\right) = F^{hn}(x) \sim G_\gamma\left(\frac{x-\mu_{hn}}{\sigma_{hn}}\right)$$

が成立する．一方で(2.1)より，

$$P\left(\max_{1\le i\le hn} X_i \le x\right) = \{F^n(x)\}^h \sim \left\{G_\gamma\left(\frac{x-\mu_n}{\sigma_n}\right)\right\}^h$$

と書けるので，

$$G_\gamma\left(\frac{x-\mu_{hn}}{\sigma_{hn}}\right) \approx \left\{G_\gamma\left(\frac{x-\mu_n}{\sigma_n}\right)\right\}^h$$

が成立することになる．この近似から

$$\mu_{hn} = \mu_n + \sigma_n\frac{h^\gamma - 1}{\gamma}, \quad \sigma_{hn} = h^\gamma\sigma_n \tag{2.2}$$

という関係式が直接計算より得られる．この(2.2)は極値に対する相似性を表現していると言える．さらに，h は自然数から実数に拡張でき，その場合は確率過程におけるブロックサイズ

の増加率に相当する．

一般化極値分布の応用のための事項をいくつか紹介する．(2.1)の確率変数を $-X_i$ に置き換えると，$P(\max_i(-X_i) \le x) = P(\min_i X_i > -x)$ であり，(2.1)において，$y = -x$, パラメータ $\alpha, \lambda > 0$ について $\gamma = -1/\alpha, \mu_n \to -\lambda, \sigma_n \to \lambda/\alpha$ とすると

$$P\left(\min_i X_i > y\right) \approx \exp\left[-\left(\frac{y}{\lambda}\right)^{\alpha}\right]$$

となり，ワイブル分布に帰着する．このとき，$\gamma < 0$ なので，(W)である．例えば，X_i を i 番目の製品の故障までの時間とすると，$P(\min_i X_i > y)$ は “製品は少なくとも y(時間)までは故障しないという事象の確率” に相当する．このように，信頼性工学において故障寿命予測によく用いられるワイブル分布（Weibull, 1939, 1951）は極値統計学に深い関連がある．同様に，極値統計論は最弱リンク理論(weakest link theory)の基礎にもなっている（Meeker et al., 2021）．

また，F を正規分布 $\mathcal{N}(0,1)$ とする．このとき，$\mu_n = \sqrt{2\log n - \log(\log n) - \log(4\pi)}$, $\sigma_n = 1/\mu_n$ と取ることで $\gamma = 0$ としたときの(2.1)が成立する（de Haan and Ferreira, 2006）．これは正規分布に従う確率変数の最大値がガンベル分布に収束することを意味するが，その収束が遅いことは有名である．そのため，penaltimate 近似の議論がある（Fisher and Tippet, 1928）．一方で，F が自由度 ν の t 分布であれば，$\gamma = 1/\nu$ となることがすぐにわかる．つまり，$\gamma > 0$ となるわけであるが，t 分布が代表的なヘビーテールとなる分布であることは周知の事実であろう．t 分布の性質から $\gamma = 1/2$ で分散が発散，$\gamma = 1$ で期待値も発散することがわかる．

このように，各分布について(2.1)が成立するかどうかを確認するのは可能なものも多いが，一般にそれなりに手間であり容易ではない．ただし，ある条件下では簡単に示すことができる．いま，元の分布関数 F は 2 階微分が連続とする．このとき，von-Mises（1936）は以下を示した：(2.1)が成立するための十分条件は

$$\lim_{y \to y^*} \frac{d}{dy} \frac{1 - F(y)}{F'(y)} = \gamma$$

または

$$\lim_{y \to y^*} \frac{\{1 - F(y)\}F''(y)}{\{F'(y)\}^2} = -\gamma - 1 \tag{2.3}$$

が成立することである．ここで，$y^* = \sup\{y : F(y) < 1\}$, F', F'' は F の 1 階，2 階微分である．この性質は一般に von-Mises condition と呼ばれており，この性質を使うと指定した F について(2.1)が成立するかどうか簡単に確認できる．なお，式(2.3)はその上の式の左辺の微分を実行しただけであるが，GEV への収束(特に，ガンベルタイプ $\gamma = 0$)に対して，より具体的なイメージが可能である．例えば，$\gamma = 0$ に収束する分布関数 F に対して，ハザードレート $F'(y)/\{1 - F(y)\}$ が定数に収束する場合(例えば指数分布が該当する)，十分大きな値である y について

$$-\frac{F''(y)}{F'(y)} = \frac{F'(y)}{1 - F(y)} \approx \text{ const}$$

が成立する．この性質はガンベルタイプを特徴づける性質を表しており，ガンベルタイプの別称を指数タイプということもあるのはこの性質に由来する（北野 他, 2025）．なお，正規分布はガンベルタイプの極値分布を持つが，そのハザードレートは単調増加し，定数に収束しない．また，フレシェタイプや逆ワイブルタイプもハザードレートは定数に収束しないが，それらの分布については(2.3)により，

$$\frac{F'(y)}{1 - F(y)} \propto \frac{F''(y)}{F'(y)}$$

が成立することを示唆している．

2.2 GEV のパラメータ推定

前章で触れたように GEV は最大値の分布であるから，データからある意味での最大値を計算しないといけない．しかし，データ全体の最大値はひとつだけであり，そこからパラメータ推定をするのは難しい．そこで，以下に述べるブロック最大化法(block maxima)を用いるのが通例である．まず，データ $X_1, \ldots, X_n$ を k 個のブロックに分割する：$\{X_1, \ldots, X_{n_1}\}$, $\{X_{n_1+1}, \ldots, X_{n_2}\}, \ldots, \{X_{n_{k-1}+1}, \ldots, X_n\}$．便宜上，$n_0 = 1$, $n_k = n$ と書く．このとき，各ブロックの最大値を $Y_i = \max_{n_{i-1}<j\leq n_i} X_j$ とすると，$\{Y_1, \ldots, Y_k\}$ が集まる．この k 個のデータに対して，GEV(μ, σ, γ) を想定し，位置，尺度，形状パラメータ (μ, σ, γ) を推定する．GEV の密度関数は

$$f_{\mathrm{GEV}}(x|\mu,\sigma,\gamma) = \frac{1}{\sigma}\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}-1}\exp\left[-\left(1+\gamma\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right]$$

なので，最尤推定量は

$$(\hat{\mu},\hat{\sigma},\hat{\gamma}) = \underset{\mu,\sigma,\gamma}{\mathrm{argmax}} \sum_{i=1}^{k} \log f_{\mathrm{GEV}}(Y_i|\mu,\sigma,\gamma)$$

と定義される．

このように，ブロック最大化法はデータをブロックに分割して疑似的に最大値を複数個構成して，それを元にパラメータ推定をする．ここで議論になるのがデータのブロック分割方法であろう．例えば，X_i が期間内で i 日目に観測されたデータとしよう．すると，$\{X_i : i = 1, \ldots, 365\}$ は最初の 1 年間で得られたデータになる．よって，数十年分の観測データがある場合，各年単位をブロックとすると年最大値データを考えることになり，それを元に作られた GEV は年最大値分布と解釈される．同様に，年最大値の代わりに月最大値をはじめとする任意の期間最大値分布を GEV でモデリングすることが可能である．ただし，注意点もある．例えば，2.5 節，3.5 節，5.3 節でも扱っている日降雨量データに対して月最大値分布を考えたいとしよう．しかし，日本では基本的に豪雨は台風シーズンの夏季に発生しやすく，冬季では発生しにくいため，各月の降雨量が同分布に従うという仮定が非現実的となる．そのため，一般に時系列構造だけでなく，解析対象のデータの領域知識を考慮したブロック分割を行う必要がある．また，時系列構造がない独立に得られたデータにおいても，ブロックの決め方は難しい．データをランダムに等分割することは可能であるが，上の例と同じく各ブロックのデータを同分布と見なしてよいか，さらには，“何の最大値を考えているのか?” といった解釈性が問題となる．GEV によるモデリングでは基本的に意味のあるブロック分割を考える必要があり，適切な分割数 k は解析の目的に応じて定まると言える．

2.3 最尤推定量の漸近理論

パラメータベクトル $\boldsymbol{\theta} = (\mu, \sigma, \gamma)^\top$ について，GEV における対数尤度関数を

$$\ell_{k,\mathrm{GEV}}(\boldsymbol{\theta}) = \ell_{k,\mathrm{GEV}}(\mu,\sigma,\gamma) = \frac{1}{k}\sum_{i=1}^{k}\log f_{\mathrm{GEV}}(Y_i|\mu,\sigma,\gamma)$$

と書く．また，フィッシャー情報行列を

$$I_{GEV}(\boldsymbol{\theta}) = E\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ell_{k,\mathrm{GEV}}(\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}^\top}\ell_{k,\mathrm{GEV}}(\boldsymbol{\theta})\right]$$

とする．このとき，Bücher and Segers (2017) は，$Y_1, \ldots, Y_k$ が独立同一に $\mathrm{GEV}(\mu, \sigma, \gamma), \gamma > -1/2$ に従うとき，最尤推定量 $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\gamma})$ について

$$\sqrt{k}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}_3(\mathbf{0}_3, I(\theta)^{-1}) \tag{2.4}$$

が成立することを示した (Smith, 1985 も参照されたい)．ここで，$\mathbf{0}_3 = (0,0,0)^\top$, $\mathcal{N}_p$ は p 変量正規分布を表す記号である．

この結果で注意したいのは，推定量の収束レートが元々のデータ数 n ではなく，元のデータから構成されたブロック数 k で決まるところにある．一般に極値理論によると，Y_i が GEV に従うのは n が大きいときの極限での話なので厳密には $Y_i \sim \mathrm{GEV}(\mu, \sigma, \gamma)$ という仮定は少々強いものとなっている．したがって，上の結果では漸近分布の期待値は 0 となっているが，本来はブロックの作り方に応じたバイアスが生じるはずである．そうでないと，収束レートが k に依存するので，結局 $k = n$ のときが最も推定量の精度が良いことになる．しかし，$k = n$ はすべてのデータを使うことを意味し，“GEV が最大値の分布である” ことに反してしまう．また，上でも述べたように，GEV による解析ではブロックはデータの領域知識から定まることが多い．その意味で，一般には (2.4) のように，$\mathrm{GEV}(\mu, \sigma, \gamma)$ をブロックとその数 k を固定としたパラメトリックモデルとして考えることになる．

2.4 GEV の再現期間 (Return Period) と再現レベル (Return Level)

極値統計モデリングを実データに適用し，その結果を検討する際に特に重要な概念となるのが再現期間 (Return Period)，再現レベル (Return Level) である．いま，$Y \sim \mathrm{GEV}(\mu, \sigma, \gamma)$ とすると，

$$P(Y > y) = 1 - \exp\left[-\left(1 + \gamma \frac{y - \mu}{\sigma}\right)^{-1/\gamma} \right]$$

より，与えられた確率 $p \in (0,1)$ に対して，$p = P(Y > z)$ を満たす $z = z(p)$ は

$$z(p) = \begin{cases} \mu + \frac{\sigma}{\gamma}\left\{ [-\log(1-p)]^{-\gamma} - 1 \right\}, & \gamma \neq 0, \\ \mu - \sigma \log(-\log(1-p)), & \gamma = 0 \end{cases}$$

で与えられる．例えば，想定規模に対するブロックサイズを T とし，$p = 1/T$ とすると，ブロック内で生成される T 個のサンプルのうち 1 つが $z(1/T)$ を超過すると解釈できる．これを書き直すと

$$T = \frac{1}{P(Y > z(1/T))} \tag{2.5}$$

である．さて，ここで Y がどのような確率変数であったかを思い出してほしい．GEV の構成によると，Y はブロック最大値であった．そして，多くの実用例では GEV は時系列で得られたデータに利用される．例えば，1 日単位で得られるデータに対してブロックサイズを 365 とすると Y は年最大値である．そのとき，T 個のサンプルのうち 1 つが $z(1/T)$ を超過するというのは，T 年間のうち 1 年間は得られるデータの最大値が $z(1/T)$ を超過すると解釈できる．言い換えると，$z(1/T)$ を超えるのは T 年間のうち (平均的に) 1 年程度であるというわけである．このように，GEV が降雨量のような時系列データの極端な事象の確率モデルに利用されていることから T は “期間” と解釈されることが多く，固定した確率 p の逆数 T とその分位点 $z(p) = z(1/T)$ はそれぞれ再現期間，再現レベルと呼ばれる．

例えば，建造物や製品の耐久性試験に関わる衝撃力データ (波の高さ，地震の大きさ，自動車の衝突時の速度など含む) では，今後 50, 100, 200 年間など長期間を想定した T に基づく $z(1/T)$ を計算し，その値に耐えうる製品を造るような設計がなされる．その意味で，T は実際

の観測期間よりも長い期間を想定されることが多い．この再現期間，再現レベルという名称は極値統計学以外ではあまり見かけないが，実際には再現レベルは分位点そのものであるので，その点に注意すると理解しやすい．

再現期間については次のような解釈もできる．ここで設定しているように，ブロックサイズを 1 年とすると，サンプルは 1 年に 1 回しか得られない．よって，想定サンプルサイズ T が $T = 1, 2, \ldots$ と増えていく状況では同じように 1 年, 2 年... と時間が経過していく．つまり，T は何らかの時間である．ここで T_0 を生起確率 p の幾何分布に従う確率変数とみると $E[T_0] = 1/p$ となることから，$T = E[T_0]$ とみるとこれは事象 $z(p)$ が起こるまでの平均的な待ち時間と解釈できる．このように再現期間 T を幾何分布の期待値であると解釈をすれば，極端な事象 $z(p)$ が T 年に 1 度は必ず起こるわけではないことを強く認識できる．

さて，再現期間・レベルは，パラメータ推定値の診断図としても機能する．まず期間 $T > 1$ に対して，$(T, z(1/T))$ を再現レベル図(Return Level Plot)と呼ぶ．データ $Y_1, \ldots, Y_k$ が GEV にうまく当てはまっているならば，$Y_1, \ldots, Y_k$ の順序統計量 $Y_{(1)} \geq \cdots \geq Y_{(k)}$ と，再現期間点(確率点の逆数)$T_i = 1/p_i, p_i = i/(k+1), i = 1, \ldots, k$ として，$\{(T_i, Y_{(i)}) : i = 1, \ldots, k\}$ の多くの点が曲線 $(T, z(1/T))$ の周りに分布することが期待できる．これはいわゆる QQ-plot の考え方と同じである．なお，$z(1/T)$ は未知パラメータ (μ, σ, γ) に依存するので，実際には推定値をプラグインする．

再現期間・レベルについて，上では元となるデータ Y_i は年最大値，つまりブロックは 1 年間であるとしていた．それでは，一般にブロックが $N(> 1)$ 年，つまり，元となるデータが N 年最大値である場合はどのように考えるのだろうか．いま，N 年最大値の確率変数を $Y^N \sim GEV(\mu_N, \sigma_N, \gamma)$ と書く．(Y^1, μ_1, σ_1) は先ほどまでの 1 年最大値 (Y, μ, σ) に対応する．このとき，GEV の分布関数を $G_{N,\gamma}((x - \mu_N)/\sigma_N)$ とし，$p^N = 1 - G_{N,\gamma}((x - \mu_N)/\sigma_N)$ を固定したときの x についての逆数を $z_N(p^N)$ とする．このとき，再現期間を T^N と書くと，

$$T^N = \frac{N}{p^N}$$

で表される．ここで，p^N, T^N はそれぞれ p, T の N 乗という意味ではなく，新しい記号として導入している．これを用いて再現レベル図を作成することは当然できるのだが，すると当然ブロックの作り方によって結果が変わってくる．このことは数学的には自然なことなのだが，データ解釈の立場に立つと理解をしづらい．ちなみに，(2.2) より再現レベルはブロックの作り方に依存しない．これを踏まえて，もう少し再現期間について考察してみる．Coles (2001) は別の再現レベル図として $(1/\{-\log(1 - p^N)\}, z(p^N))$ を用いている．これは再現期間を改めて

$$T^N = \frac{N}{-\log(1 - p^N)} \tag{2.6}$$

と定義していることと同値である．実はこの設定の下で，(2.2) より (2.6) は任意の N で

$$T = T^N = \left(1 + \frac{\gamma(x - \mu)}{\sigma}\right)^{1/\gamma}$$

となり，ブロックの作り方に依存しなくなることがわかる．したがって，この新しい T を用いて $(T, z(1/T))$ を再現レベル図として定義する方がデータ解釈の意味では理解しやすい．これを踏まえると再現期間の定義は 2 通りあることになり，Tawn (1988) もこの点に言及している．また，北野 (2020) はこのことについて，稀な事象の生起をポアソン分布で表現した場合のモデリングとの関連も示している．

2.5 GEV の適用例

1976 年 1 月 1 日〜2023 年 12 月 31 日までに東京都八王子市の降雨量観測所で観測された日降雨量データに一般化極値分布を適用する．降雨量を始めとする気象関連データは気象庁の HP からダウンロードできる．八王子市観測所は 1976 年 1 月 1 日から降雨量観測を始めており，1975 年以前はデータの蓄積がない．さて，48 年間でデータの蓄積は 17521 日分あるが，そのうち降雨がない日(降雨量 0)は 12097 日で 69% に相当し，実際に降雨があった日は 5424 日であった．降雨日は年平均 113 日(標準偏差 12.2) であった．図 1 に降雨量をプロットしている．図中の赤点は各年の最大降雨量である．

さて，GEV は独立同分布に従う確率変数に対して適用される．よって，まずは独立同分布性に言及する．降雨量データが各日で同分布に従うことを調べるのは困難であるが，興味あるのは分布の裾部分なので全体を表現する分布には興味がない．しかし，例えばブロックサイズを 1 年とした年間の最大降雨量に同分布を仮定するのは自然であろう．次に独立性に言及する．降雨量データは通常，データ全体が独立に生起しているとは考えにくい．しかし，今回は豪雨のみに関心があり，晴れや小雨には関心がない．よって，極値事象に限定した自己相関のような従属性の有無を調べる．ここでは，Coles et al. (1999)が提案した(上側)裾従属係数(裾依存係数とも言う)を利用した

$$\chi(\ell) = \lim_{q\to 1} P(F(X_{i+\ell}) > q | F(X_i) > q) = \lim_{q\to 1} \frac{P(F(X_{i+\ell}) > q, F(X_i) > q)}{1-q}, \quad \ell \geq 1$$

を用いる．Spearing et al. (2023)もこの基準を用いて極値事象の時系列の定常性の有無を調べている．裾従属係数は本来 2 変量の閾値超過確率であり，次章で述べる単変量の閾値超過確率である一般化パレート分布と密接に関連している．簡単にその性質を述べると，$0 \leq \chi(\ell) \leq 1$ であり，$\chi(\ell) = 0$ のときは裾独立(asymptotic independence)，$\chi(\ell) > 0$ のとき，裾従属(asymptotic dependence)と言われる．また，$\chi(\ell) = 1$ のときは完全従属(full dependence)である．裾従属係数の計算は統計ソフト R のパッケージ extRemes が便利である (Gilleland and Katz, 2016)．ここでの裾従属係数 $\chi(\ell)$ は，時系列に得られたデータについて i 日目の(一様分布に変換された)データが大きな値を持ったとき，続けて ℓ 日後にもまた大きな値を持つ確率と解釈される．特に，降雨量データに対して $\ell = 1$ とすると，ある程度の大雨が降った日の翌日も大雨であった確率を見ていることになる．これは “2 日連続で大雨が降る確率” とは異なるので注意して解釈されたい．実際には $\mathcal{X}(\ell)$ は 1 に近い q で，F や P は経験分布関数で近似する．八王子市の日降雨量では，全体の 98% 分位点が 46mm に相当した．よって，$q = 0.98$ で計算すると，$\mathcal{X}(\ell) = 0.131$ であった．つまり，観測期間全体を通してある程度の降雨があった日の翌日も同程度かそれ以上の降雨であるケースは稀であったので，降雨量に関する極値事象は独立に生起されると仮定して GEV を当てはめる．ちなみに $\ell > 1$ とすると裾従属係数はより小さい値に

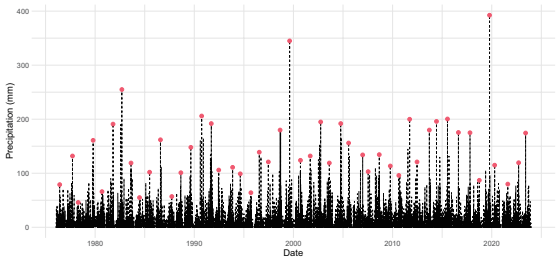

図 1. 1976 年 1 月 1 日〜2023 年 12 月 31 日における八王子市での日降雨量(mm)．赤点は各年(48 年分)の最大降雨量．

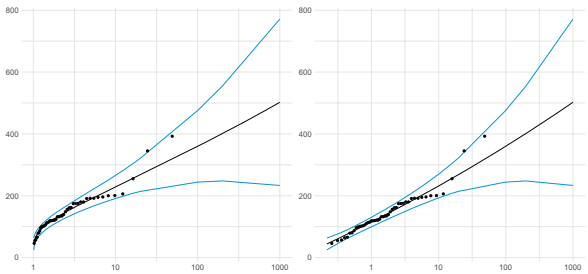

図 2. 八王子市の降雨量データの再現期間(x 軸)と再現レベル(y 軸)．左は再現期間を $T = 1/p$, 右は再現期間を $T = 1/(-\log(1-p))$ としたもの．

なったが，これは直感的にも明らかであろう．気象データに関する推測では季節依存なども重要な要素となる．特に降雨量は台風シーズンは大きな値を取る傾向がある．しかし，今回はそのような解析は省略する．

さて，八王子市の降雨量データの極値解析に移る．ここではブロックサイズは 1 年とする．つまり，年最大値データ(図 1 の赤点)を GEV に適用する．なお，解析には R パッケージ ismev を用いた (Stephenson, 2018)．すると，パラメータの最尤推定値は $(\hat{\mu}, \hat{\sigma}, \hat{\gamma}) = (115, 48.1, 0.04)$ となった．形状パラメータ γ の 95% 信頼区間は $[-0.156, 0.236]$ となった．形状パラメータは符号の正負どちらも取り得るため，それほどヘビーテールしていないことがわかる．言い換えると，年最大値データはヘビーテールする可能性はあるもののそれなりに安定した裾挙動をしていることがわかる．続いて再現期間・レベルを確認する．図 2 には 2 つの再現期間に対する再現レベルをプロットしている．図中の曲線(黒)が再現期間に対する再現レベル，点が年最大値データの再現期間に対する QQ-plot, 青線が漸近理論から求められる再現レベルの 95% 信頼区間である．左は再現期間が $T = 1/p = 1/P(X > z(1/T))$ なので $T > 1$ である．右図は再現期間 $T = 1/\{-\log(1 - P(X > z(1/T))\}$ なので，$T > 0$ が定義域となる．このブロック最大によって得られた結果は，1 年最大値データであることを強く意識するのであれば左図が理解しやすい．しかし，それゆえにブロックサイズの柔軟性はない．他方，右図は(2.2)に示した相似性の図示表現となっているため，ブロックサイズによらない解釈ができる．このモデリングは観測期間 48 年としたため，$T > 48$ のデータが存在しない．しかし，再現レベル図から $T > 48$ の将来の豪雨時の降雨量を見積もることができる．

GEV を用いた波浪や潮位などに関する気象データ解析は Haigh et al. (2010)，Caires (2011a, 2011b)，Wahl et al. (2013) なども参照されたい．

2.6 上位順序統計量の極値分布

GEV は観測データの最大値分布として定義されていた．パラメータ推定の際はブロック最大値を利用したが，それでも推定に使うデータ数は少ない．また，あるブロックの 2 番目に大きいデータは他のブロックの最大値よりも大きな値を持つこともあるだろう．そこで，各ブロックについて上位 1 番目から r 番目 $(r > 1)$ までのデータを使ってパラメータ推定を行う方法もある (Smith, 1986; Tawn, 1988)．いま，$X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$ の順序統計量を $X_{(1)} \geq \cdots \geq X_{(n)}$ とする．このとき，F が(2.1)を満たすとすると，$(X_{(1)}, \ldots, X_{(r)})$ の同時分布は $n \to \infty$ の下で密度関数

$$g(x_{(1)}, \ldots, x_{(r)}) = \exp\left[-\left[1+\gamma\left(\frac{x_{(r)}-\mu}{\sigma}\right)\right]^{-1/\gamma}\right] \prod_{i=1}^{r} \frac{1}{\sigma}\left[1+\gamma\left(\frac{x_{(i)}-\mu}{\sigma}\right)\right]^{-1/\gamma-1} \tag{2.7}$$

を持つ極値分布で近似される．ここで，$x_{(1)} \geq \cdots \geq x_{(r)}$ であり，$1+\gamma(x_{(i)}-\mu)/\sigma > 0 (i = 1, \ldots, r)$ である．$\gamma = 0$ のときの密度関数は(2.7)を $\gamma \to 0$ として得られる．ここで $r = 1$ のときは(2.1)に帰着する．(2.7)の証明は高橋・志村 (2016)の 2.3 節が詳しい．

パラメータ (μ, σ, γ) の推定の際は，各ブロックの $1 \sim r$ 番目のデータが独立に(2.7)に従うと見なして最尤推定を行えばよい．ブロック数を k とするとパラメータ推定に使えるデータ数が kr 個となり推定値が安定する．Guades Soares and Scotto (2004)は検定に基づく r の決め方を議論している．

3. 一般化パレート分布

3.1 一般化パレート分布(Generalized Pareto distribution)

さて，(2.1)が成立するとき，元の分布関数 F は閾値 w について，w に依存する尺度 σ_w が存在して

$$P(X_i > w + \sigma_w x | X_i > w) = \frac{P(X_i > w + \sigma_w x)}{P(X_i > w)} \to (1+\gamma x)_+^{-1/\gamma} \quad \text{as } w \to \infty \tag{3.1}$$

が成立する(証明は付録 A を参照)．これを書き直すと，改めて $\sigma = \sigma_w$ とおいて

$$P(X_i - w > x | X_i > w) \approx \left(1 + \frac{\gamma x}{\sigma}\right)_+^{-1/\gamma}$$

を得る．右辺の分布は一般化パレート分布(Generalized Pareto distribution, GP)と呼ばれ，尺度パラメータ σ, 形状パラメータ γ を用いて $GP(\sigma, \gamma)$ で表現される．$\gamma = 0$ のとき，右辺は $1 - e^{-x/\sigma}$ と指数分布になり，$\gamma < 0$ で有限のエンドポイント $x < -\sigma/\gamma$ が得られる．GP はデータ全体のうち閾値 w より大きい部分が持つ確率分布であると言える．また，γ の符号に関する解釈は GEV と同じである．形状パラメータは閾値に依存せず，元の分布関数 F にのみ依存するが，尺度パラメータは閾値 w と F に依存する．しかし，結局 F が未知なので，σ_w 自身をパラメータだと思ってこれも推定するのが基本方針となる．この方法の枠組みは，閾値を超える部分のモデリングすることから，閾値超過法(peak over threshold method)と呼ばれている(Davison and Smith, 1990)．

3.2 GP の最尤推定

GP では閾値より大きいデータのみで解析する．いま，閾値 w に対して，$X_i > w$ となるデータについて $Y_i = X_i - w$ とし，$\{Y_1, \ldots, Y_k\}$ と再ラベリングする．この k 個のデータに GP を当てはめる．GP の密度関数は

$$f_{\mathrm{GP}}(y|\sigma, \gamma) = \frac{1}{\sigma}\left(1 + \frac{\gamma y}{\sigma}\right)_+^{-1/\gamma - 1}, \quad y > 0$$

であるので，最尤推定量は

$$(\hat{\sigma}, \hat{\gamma}) = \underset{\sigma, \gamma}{\mathrm{argmax}} \sum_{i=1}^{k} \log f_{\mathrm{GP}}(Y_i | \sigma, \gamma)$$

である．

統計モデル $\mathrm{GP}(\hat{\sigma}, \hat{\gamma})$ はその分位点を用いて診断できる．いま，固定した $p \in (0, 1)$ に対して，$p = P(Y_i > y)$ を満たす $y = y(p)$ は

$$y(p) = \frac{\sigma}{\gamma}(p^{-\gamma} - 1)$$

となる．よって，$p_i = i/(k+1), i = 1, \ldots, k$, $\{Y_1, \ldots, Y_k\}$ の順序統計量を $Y_{(1)} \geq \cdots \geq Y_{(k)}$ とすると，モデルの当てはまりが良ければ $\{(y(p_i), Y_{(i)}) : i = 1, \ldots, k\}$ は 45 度線 $y = x$ に沿うように分布するはずである．実際には，(σ, γ) にはその推定量 $(\hat{\sigma}, \hat{\gamma})$ を代入する．これは実質的に GP に対する QQ-plot である．

GP は多くのデータに適用可能である．しかし，GP は GEV とは違いデータ全体に対して閾値を設定するので，年最大などの閾値の取り方で分布の解釈を与えるのは難しい．したがって，閾値 w, または k の選択が重要となる．GEV のようにデータ解釈ベースに設定されることもあるが，基本的には調整パラメータと思って何らかの意味で最適化する必要がある．閾値の選択法は様々なものが提案されているが，ここでは標準的な方法である平均残差生存図 (mean residual life plot, Coles, 2001) を紹介する．平均残差生存図は平均超過量図 (mean excess plot) とも呼ばれる．ここでは，$\gamma < 1$ を仮定する．これは実用上は決して強い仮定ではない．実際，$\gamma = 1$ はコーシー分布の裾と同程度であり，かなり重い裾である．閾値 w_0 と確率変数 $X - w_0 \sim \mathrm{GP}(\sigma, \gamma), (X > w_0)$ について

$$E[X - w_0 | X > w_0] = \frac{\sigma_{w_0}}{1 - \gamma}$$

が成立する ($\gamma \geq 1$ のとき，期待値は発散)．ここで，$w > w_0$ となる別の閾値について

$$E[X - w | X > w] = \frac{\sigma_w}{1 - \gamma} = \frac{\sigma_{w_0} + \gamma w}{1 - \gamma}$$

となり (付録 A を参照)，条件付き期待値は w_0 より大きい閾値 w に関して傾きが固定の線形構造を持つ．よって，$k = k_w = \sum_{i=1}^{n} I(X_i > w)$ とするとき，

$$E[X - w | X > w] \approx \frac{1}{k} \sum_{i=1}^{n} (X_i - w) I(X_i > w)$$

となるので，$w > w_0$ の部分で

$$\left(w, \frac{1}{k} \sum_{i=1}^{n} (X_i - w) I(X_i > w) \right) \tag{3.2}$$

は線形構造を持つはずである．このように，(3.2) を図示して，安定的に線形構造を有する部分として w_0 を決定すればよいことがわかる．ただし，極値統計論としては $E[X - w | X > w]$ の w に関する線形構造に着目するのは妥当であるが，実データに対する (3.2) は簡単に解釈できないことが多い (明確な線形構造を見出すことが難しい)．特に，w が大きくなりすぎると，$X_i > w$ となるデータ数が少なくなり，不安定な挙動をすることが多くなる．実際には，(3.2) により判断できるのは，適切な閾値が存在すると考えられる区間 (これを (w_L, w_U) と表す) にとどまり，その区間内で実際にパラメータ推定を行い，推定値の挙動が安定している箇所を最適値として用いる．なお，区間 (w_L, w_U) の設定についても，(3.2) のグラフの形状やデータ知識などをもとに解析者の判断により決定される．パラメータの推定値の挙動であるが，極値統計論によると γ は閾値に対して不変であるが，尺度パラメータは $w > w_L$ について，$\sigma_w = \sigma_{w_L} + \gamma(w - w_L)$ となり，閾値に対して線形となる．よって，

$$\sigma^* = \sigma_w - \gamma w \equiv \mathrm{const}$$

とすれば σ^* は理論的には閾値不変となる．このように，閾値 (w_L, w_U) の範囲で $(w, \hat{\sigma}^*)$, $(w, \hat{\gamma})$ の安定性を評価し (定数的挙動をしていることを確認)，最終的に適切と思われる閾値をひとつ

決定する．なお，閾値 w が決まれば，閾値より大きいデータ数 $k=\sum_{i=1}^{n} I(X_i>w)$ も決まる．

上記の方法は閾値をオートマティックに決定するものではない．伝統的なオートマティックな閾値選択方法はレビュー論文 Scarrott and MacDonald (2012) を参照されたい．近年では Wadsworth (2016)，Northrop et al. (2017)，Murphy et al. (2024) などが新規手法を開発している．

3.3 GP の漸近理論

GP では閾値を調整パラメータと見ることが多いため，その影響を考慮した理論構築がなされる．そこで (3.1) を改めて

$$P(X_i>w+\sigma_w x|X_i>w)\approx(1+\gamma x)_+^{-1/\gamma}+A(w)B(x)\ \text{ as } w\to\infty \tag{3.3}$$

と書く．ここで，A は $\lim_{w\to} A(w)=0$ を満たす関数であり，$B(x)$ は x の関数である．A や B の具体的な形は de Haan and Ferreira (2006) の Theorem 2.3.8 で与えられている．要は $A(w)$ が閾値の選択による GP の近似バイアスであり，これが推定量 $(\hat\sigma,\hat\gamma)$ のバイアスに影響するのである．

GP の対数尤度関数を改めて

$$\ell_{k,\mathrm{GP}}(\sigma,\gamma|w)=\frac{1}{k}\sum_{i=1}^{k}\log f_{\mathrm{GP}}(X_i-w|\sigma,\gamma)I(X_i>w)$$

と書く．GP における最尤推定量 $(\hat\sigma,\hat\gamma)$ の漸近理論は 2 通りのものが知られている．ひとつは Smith (1987) によるもので，この論文では閾値 $w=w_n$ は n に応じて変動する実数列と考えて，$n\to\infty$ のとき，$w_n\to\infty$, $P(Y_i>w_n)\to 0$, $nP(Y_i>w_n)\to\infty$ と仮定している．つまり，閾値より大きいデータの個数 $k=\sum_{i=1}^{n} I(X_i>w_n)$ 自体が確率変数で $k/n-P(Y_i>w_n)\xrightarrow{P}0$ である．すると，$P(Y_i>w_n)\to 0$ という仮定は実質的に $k/n\to 0$ と同じであり，元のデータ数 n が増える度に閾値より大きいデータ数 k も増えるが n と比較するとその数は少ないという状況を表現していることになる．さて，このような状況で，また適切な仮定の下で

$$\sqrt{k}\begin{bmatrix}\frac{\hat\sigma}{\sigma}-1\\ \hat\gamma-\gamma\end{bmatrix}\xrightarrow{D}\mathcal{N}_2(\boldsymbol{\lambda}_1,\Sigma_1)$$

が成り立つ．ここで．$\boldsymbol{\lambda}_1=(\lambda_{1,1},\lambda_{1,2})$ は $A(w_n)$ や $B(x)$ に依存して決まるバイアス項であり，$A(w_n)$ と k によって決まる関数 $\boldsymbol{\lambda}(k)$ に対して，$\sqrt{k}\boldsymbol{\lambda}(k)\to\boldsymbol{\lambda}_1$ と仮定されていると解釈できる．このバイアスの詳細な式は Smith (1987) を参照されたい．また，Σ_1 は

$$\Sigma_1=\begin{bmatrix}2(\gamma+1) & -(1+\gamma)\\ -2(1+\gamma) & (1+\gamma)^2\end{bmatrix}$$

であり，これは $\ell_{k,GP}(\sigma,\gamma|w)$ のフィッシャー情報行列から導かれる．一般的に，w_n が大きいとデータは GP によく当てはまるので (3.3) の近似精度が上がる ($A(w_n)$ が小さくなる)．しかし，その分 k が小さくなり，パラメータ推定に用いるデータ数が少なくなることから推定量の分散は増大する．逆に，w_n が小さいと k は大きくなり推定量の挙動は安定する (分散が減少) が，(3.3) の精度が悪くなる ($A(w_n)$ の値が大きくなる)．そのため，収束レートである $k^{-1/2}$ はバイアスと分散のトレードオフのバランスを取る役割を果たしていることになる．

Drees et al. (2004) は Smith (1987) とは別のシナリオを考えている．元のデータ $X_1,\ldots,X_n$ の順序統計量を $X_{(1)}\geq\cdots\geq X_{(n)}$ とするとき，彼らは閾値を上から $k+1$ 番目のデータ：$w=X_{(k+1)}$ と仮定している．つまり，パラメータ推定に使うデータは $\{Y_i=X_{(i)}-X_{(k+1)}, i=1,\ldots,k\}$ で

あるとみなせる．その上で，Smith（1987）と同じ解釈が可能な設定：$n \to \infty$ の下で $k \to \infty$, $k/n \to 0$ を考えている．大まかにいうと，Smith（1987）の設定では閾値 w が実数，閾値を超えるデータ数 k が確率変数であり，Drees et al.（2004）では閾値が確率変数，閾値を超えるデータ数 k は実数となっている．この設定は，2.6 節で述べた上位順序統計量の極値分布においてブロック数を 1 とした場合に関連している．Drees et al.（2004）は以上のような設定といくつかの適切な条件の下で

$$\sqrt{k}\begin{bmatrix} \frac{\hat{\sigma}}{\sigma} - 1 \\ \hat{\gamma} - \gamma \end{bmatrix} \xrightarrow{D} \mathcal{N}_2(\boldsymbol{\lambda}_2, \Sigma_2) \tag{3.4}$$

が成立することを示した．ただし，$\boldsymbol{\lambda}_2 = (\lambda_{21}, \lambda_{22})$ は Smith（1987）の理論と同じく分布の近似バイアス $A(w)$ に依存する項である．しかし，$A(w)$ の設定が Smith（1987）とは異なるので $\boldsymbol{\lambda}_1$ との直接の比較は難しい．また，分散共分散行列は

$$\Sigma_2 = \begin{bmatrix} 2(1+\gamma)+\gamma^2 & -(1+\gamma) \\ -2(1+\gamma) & (1+\gamma)^2 \end{bmatrix}$$

で与えられる．ここで注目するのは Σ_2 の $(1,1)$ 成分が Σ_1 のそれとは異なっていることである．この違いは閾値の取り方の違いによるものである（本質的な意味合いはほとんど等しいとしても）．結果から閾値を実数とした方が尺度パラメータ $\hat{\sigma}/\sigma$ の推定量の分散が小さいことがわかる．ただし，推定量の分散が小さいからと言って閾値を固定で考える方がよいと結論付けるのは早計である．なぜならば，両者のバイアスは一概にどちらが小さいか判断できるものではないためである．またバイアスは，要は GP の近似バイアスから発生しているが，その分布関数の近似誤差の推定は一般に困難である．分布関数の近似誤差，ひいてはバイアスを推定する研究も Gomes and Martins（2002），Gomes and Pestana（2007），Caeiro and Gomes（2011）などで議論されているが，このバイアスを補正した推定量を考える場合の最適な閾値 w（または閾値より大きいデータ数 k）は補正しない場合とは異なるのでやはり直接の比較は難しい．

このような背景から，(σ, γ) の信頼区間や仮説検定を行う場合はバイアスを 0 と仮定し，Σ_1, Σ_2 に γ の推定量を代入するいわゆるプラグイン法が主流となっている．近年では，de Haan and Zhou（2024）が極値解析におけるブートストラップ法の理論的正当性を示しているため，ブートストラップ信頼区間などは今後広く利用されると思われる．

なお，Drees et al.（2004）の理論は推定に用いるデータが $Y_i = X_{(i)} - X_{(k+1)}$ なので $Y_1, \ldots, Y_k$ が独立な確率変数ではない．したがって，(3.4) は通常の最尤推定の証明が通用せず，順序統計量，極値統計理論，確率過程を融合させた証明を元に構築されている．

3.4 GP の再現レベル

いま，$X \sim F$ として，閾値 w を超えた部分が $\mathrm{GP}(\sigma, \gamma)$ に従うとすると，$x > w$ に対して，

$$\begin{aligned} P(X > x) &= P(X > w)P(X > w + x - w | X > w) \\ &= P(X > w)\left(1 + \frac{\gamma(x-w)}{\sigma}\right)^{-1/\gamma} \end{aligned}$$

と表すことができる．したがって，固定した $p \in (0,1)$ に対して，$p = P(X > z)$ を満たす $z = z(p)$ は

$$z(p) = w + \frac{\sigma}{\gamma}\left\{\left(\frac{p}{P(X>w)}\right)^{-\gamma} - 1\right\}$$

となる．X と同じ分布を持つ元のデータ $X_1, \ldots, X_n$ に対して閾値 w を超えるデータ数を k と

すると，$P(X > w) \approx k/n$ と推定できるので，これに置き換えて

$$z(p) \approx w + \frac{\sigma}{\gamma}\left\{\left(\frac{np}{k}\right)^{-\gamma} - 1\right\}$$

と近似される．実際には，(σ, γ) にはその推定量 $(\hat{\sigma}, \hat{\gamma})$ が用いられる．

GP に対しても再現レベルという言葉は使われる．想定サンプルサイズ m に対して，$p = 1/m$ とすると，m 回に 1 回程度 $z(p) = z(1/m)$ を超過するイベントが発生すると解釈できる．また，もしデータが 1 年で R 個で得られたとすると ($R = 365$ とすると 1 日単位で得られるデータに相当)，$T = m/R$ として $z(1/m) = z(1/(RT))$ は T 年に 1 回超過する値，つまり再現期間 T 年の再現レベルであると解釈できる．想定サンプルサイズの考え方が GEV と若干異なるので注意されたい．GP ではこの T が(年数の)再現期間と定義される．これをもう少し紐解くと，$1/p = m = RT$ より，

$$T = \frac{1}{Rp} = \frac{1}{RP(X > z(1/T))} = \frac{1}{RP(X > w)}\left(1 + \frac{\gamma(z(1/T) - w)}{\sigma}\right)^{1/\gamma}$$

となる．

再現レベルの定義は GEV に対応させて

$$\frac{1}{m} = 1 - \exp[-P(X > x)] \approx 1 - \exp\left[-\left(1 + \frac{\gamma(x - w)}{\sigma}\right)^{-1/\gamma}\frac{k}{n}\right]$$

を満たす x として定義されることもある．この場合，x は $P(X > x) = -\log(1 - 1/m)$ の逆数として計算される．m が大きいときは $\log(1 - 1/m) \approx -1/m$ なので上のケースとほぼ等しい．3.5 節の実例では年数再現期間 T について $m = 365 \times T$ を想定したが，どちらも違いはほとんどなかった．

2.4 節では GEV の再現期間の定義は 2 通りあると述べたが，そのうち GP の再現期間に対応するのは(2.6)であることがその定義からわかる．

3.5 GP の適用例

2.5 節と同じ八王子市の降雨量データに今度は GP を当てはめる．GEV で利用した R パッケージ ismev でも GP を実行できるが，ここではもうひとつの極値解析の標準的なパッケージである eva を使う（Barder and Yan, 2020）．まずは閾値の選択を行う．図 3 には 3.2 節で議論した平均残差生存図(MRL，左上)とパラメータ $(w, \hat{\gamma}), (w, \hat{\sigma}^*)$ の挙動(右上，左下)を示している．これらは挙動が安定している箇所の確認が目的なので信頼区間はあえて載せていない．これらの 3 つのグラフでは縦線で (w_L, w_U) を示している．この範囲内で MRL とパラメータは比較的安定して定数挙動していることがわかる．ここでは，図中の赤丸である $w = 92$ 用いて GP を適用する．右下図にはデータと閾値を図示している．このとき，閾値より大きいデータ数は $k = 79$ で $k/n \approx 0.005$ であった．この最適な閾値を用いたときの GP のパラメータの最尤推定値は $(\hat{\sigma}, \hat{\gamma}) = (45.9, 0.05)$ となった．バイアスを 0 とした場合の形状パラメータ γ の 95% 信頼区間は $[-0.183, 0.287]$ であった．2.5 節で示した GEV の形状パラメータの推定値とほぼ等しくなっている．$\hat{\gamma}$ はほぼ 0 でありガンベルタイプに近いが，GEV 解析と同様に若干ヘビーテールとなっている．

次に，当てはまったモデル GP$(\hat{\sigma}, \hat{\gamma})$ の適合度を視覚的に確認する．図 4 の左図は 3.2 節で解説した $\{(y(p_i), Y_{(i)}) : i = 1, \ldots, k\}$ と $y = x$ の直線，また，$(\hat{\sigma}, \hat{\gamma})$ の漸近理論とデルタ法から導かれる $y(p)$ の 95% 信頼区間である．すべてのデータが信頼区間内に収まっていることが確認できる．図 4 の右図は 3.4 節において $R = 365$ と置いた再現期間(年数)T(対数スケール)と再

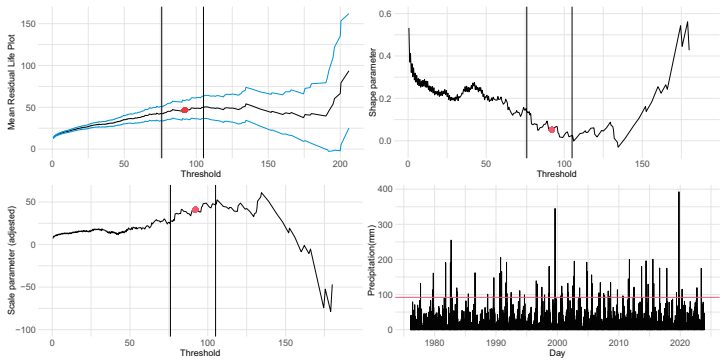

図 3. 八王子市の降雨量データの閾値選択．左上図は平均残差生存図（青線は信頼区間）．右上は $(w, \hat{\gamma})$．左下は $(w, \hat{\sigma}^*)$．赤点は $w_o pt = 92$．縦実線は (w_L, w_U)．右下図は降雨量データと閾値（赤線）．

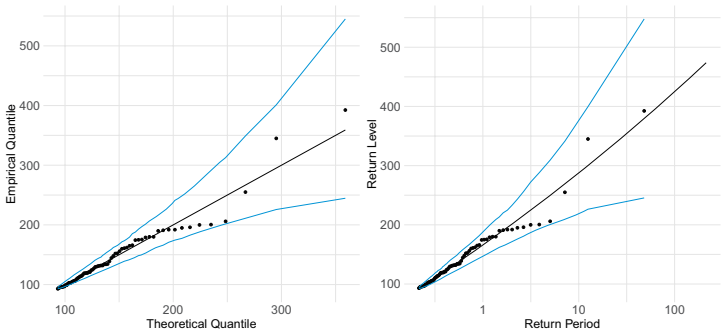

図 4. 八王子市の降雨量データに対する GP の QQ-plot（左図）と再現レベル図（右図）．

現レベル $z(1/T)$ を示している．青線は 95% 信頼区間である．1976 年以降，観測期間 48 年としたときの八王子観測所の降雨量の最大値は 392.5mm であったが，これは再現期間 50 年の再現レベルより少し大きな値となっていることが確認できる．

今回の GP を用いた解析では観測期間全体を通して日降雨量を独立な事象として扱った．しかし，日本特有の事情として，まず豪雨は台風も多い夏季に起こりやすい．よって，季節変動を考慮した深化研究は今後の解析として興味深い．

図 5 に各年の月日に対する降雨量を重ねて描いている．この図より，閾値超過する豪雨イベントは多くが夏季に集中していることがわかる．実際に閾値 w（横線）を超過しているのはほとんどが 7 月から 10 月である．このように，日本の降雨量データは本来季節性があるので，定常な GP を適用しにくく，本来は季節性を共変量として考慮するなどの工夫をするのが自然である．一方で，降雨の極値解析に伝統的に GEV, すなわち年最大値を用いるのは季節性を暗に消すためであるとも考えられる．

先に挙げた Caires (2011a, 2011b)，Wahl et al. (2013) は GEV だけでなく GP による実データ解析も行っており，このことは両方のモデルの振る舞いから総合的にデータ解釈を行うこと重要性を物語っていると言えるだろう．

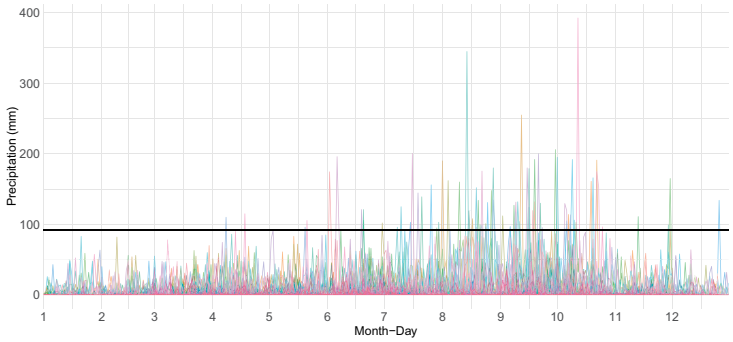

図 5. 各年の各月日の降雨量．横線は閾値 $w_{opt} = 92$.

3.6 パレート型分布

一般化パレート分布において γ の符号が正であるとわかっているとき，(3.1)において，$\sigma_w \approx w\gamma$ ととることができる．このとき，任意の $x \geq 1$ について，

$$P\left(\frac{X_i}{w} > x \middle| X_i > w\right) = \frac{P\left(\frac{X_i}{w} > x\right)}{P(X_i > w)} \to x^{-1/\gamma} \quad \text{as } w \to \infty$$

を得る．(3.5)の右辺の分布はパレート型分布(Pareto-type distribution)と呼ばれる．

一般にデータ解析をするとき，γ の符号そのものも未知であることが多いが，“データはヘビーテールする”，または “ヘビーテールであると想定しておく方が無難である” という事前情報があるならば，最初から $\gamma > 0$ と仮定し，パレート型分布を当てはめることもある．他にも例えば尖度が正規分布の尖度である 3 に比べて大きい場合はヘビーテールするとみなすこともある．特に株価のデータや(地域と種類によるが)様々な気象データは $\gamma > 0$ と仮定できる場合が多い．

GP と同様に，PT でも閾値より大きいデータのみを用いて解析する．閾値 w に対して，$X_i > w$ となるデータについて $Y_i = X_i/w$ とし，$\{Y_1, \ldots, Y_k\}$ と再ラベリングする．すると，$Y_i, i = 1, \ldots, k$ はパレート型分布(PT)に従うとみなせる．PT の密度関数は

$$f_{\mathrm{PT}}(x|\gamma) = \frac{1}{\gamma} x^{-1/\gamma - 1}, \quad x \geq 1$$

となるので，γ の最尤推定量は

$$\hat{\gamma} = \operatorname*{argmax}_{\gamma} \sum_{i=1}^{k} \log f_{\mathrm{PT}}(Y_i|\gamma) = \frac{1}{k} \sum_{i=1}^{k} \log Y_i$$

と単純な対数変換後の標本平均の形で与えられる．ちなみに，元のデータ $X_1, \ldots, X_n$ の順序統計量を $X_{(1)} \geq \cdots \geq X_{(n)}$ とし，$w = X_{(k+1)}$(大きい方から数えて $k+1$ 番目のデータ)とすると

$$\hat{\gamma} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{X_{(i)}}{X_{(k+1)}}$$

となり，最尤法とは別のアプローチで得られる Hill 推定量 (Hill, 1975)と呼ばれる推定量と一致する．

GP では閾値を固定点と見るか確率変数と見るかで $\hat{\sigma}$ の分散が異なったが $\hat{\gamma}$ は同じであった．PT ではパラメータは形状パラメータ γ のみなので，閾値の考え方の違いは漸近理論には

表れない．この場合，$\hat{\gamma}$ の漸近的性質は Hill (1975)が示しており，GP とほぼ同様の条件下で

$$\sqrt{k}(\hat{\gamma}-\hat{\gamma}) \xrightarrow{D} N(\lambda,\gamma^2)$$

が成立する．ここで，λ はやはり(3.3)の近似から発生するバイアスである．

PT を用いる場合の再現レベルの解釈は GP と同様である．固定した $p\in(0,1)$ に対して，$p=P(X>z(p))=P(X/w>z(p)/w|X>w)P(X>w)$ を満たす $z(p)$ は

$$z(p)\approx w\left\{\frac{p}{P(X>w)}\right\}^{-\gamma}$$

で与えられる．よって，$P(X>w)\approx k/n$ で置き換えると $z(p)\approx w\{(np)/k\}^{-\gamma}$ となる．3.4 節と同様に，想定サンプルサイズ T に対する再現レベル $z(1/T)$ が導かれる．閾値より大きいデータ数を k とすると，形式的に $w=z(k/n)$ となるので

$$z(p)=z(k/n)\left(\frac{p}{k/n}\right)^{-\gamma} \tag{3.5}$$

を得る．特にこの場合，より高位の分位点 $z(p)$ はそれなりに良い推定が可能な分位点 $z(k/n)$ の定数倍であるという解釈ができる．これは 4.2 節で述べる極値分位点の基礎理論にもなっている．

4. 発展

第 3 章までで極値統計学の基本的な分布である GEV, GP を用いた分析方法を述べた．本章では，それらのさらなる展開を 2 つにしぼって紹介する．ただしそれらは別々の問題というわけではなく，密接に繋がりがある．

4.1 回帰分析

興味ある変数の極値挙動を GEV, GP, PT でモデリングする方法を述べたが，そのひとつの応用が共変量を取り込んだ回帰分析である．すなわち，目的変数 Y, 共変量 X のペアを考え，共変量 $X=x$ と固定した下での Y の分布関数 $P(Y\le y|X=x)=F(y|x)$ が(2.1)を満たすと仮定する．回帰分析においては，(2.3)に対応する

$$\lim_{y\to y^*(x)}\frac{\{1-F(y|x)\}F''(y|x)}{\{F'(y|x)\}^2}=-\gamma(x)-1$$

を仮定する場合が多い．ここで，$y^*(x)=\sup\{y|F(y|x)<1\}$ である．さらに $\gamma(x)$ を x に関する関数として推定する場合は γ が x について連続微分可能である仮定を追加する．GEV の場合は 3 つのパラメータが x に依存した $\mathrm{GEV}(\mu(x),\sigma(x),\gamma(x))$ を考え，GP では $\mathrm{GP}(\sigma(x),\gamma(x))$, PT では $\mathrm{PT}(\gamma(x))$ とし，それぞれの関数の推定問題に帰着する．GEV 回帰の主な手法は，Yee and Stephenson (2007)，Castro-Camino et al. (2022)，Zhong et al. (2022)，GP 回帰は Beirlant and Goegebeur (2004)，Chavez-Demoullin and Davison (2005)，Youngman (2019)，PT 回帰は Wang and Tsai (2009)，Goegebeur et al. (2015)，Lin et al. (2022)などが挙げられる．しかし，GEV や GP では形状パラメータの推定量の精度が悪くなる例も多いことが知られており，この部分だけは定数：$\gamma(x)\equiv\gamma$ を仮定するケースも多い．また，非線形関数 $\gamma_0(x)$ に対して $\gamma(x)=\exp[\gamma_0(x)]$ と仮定し，$\gamma(x)>0$ に限定して議論させている場合も多い．同様に $\gamma(x)<0$ に限定して議論することも可能だが，それならば $\gamma(x)\equiv 0$ の単に指数分布を考えればconservative なモデルとなるので実用例は少ない．実際に，GEV や GP は形状パラメータの符号によって分布関数が(不連続に)異なるので，$\gamma(x)$ と x に依存させて関数推定した $\hat{\gamma}(x)$ が x

によって符号が変わると解釈が難しくなる．よって，理論研究でも $\gamma(x)$ の符号は固定で語られる場合がほとんどである．共変量 x によって $\gamma(x)$ の符号が変わるモデルに関する統計理論研究は未だに未開発であり，重要な課題である．極値回帰の R パッケージとしては Youngman (2022) が evgam を開発している．このパッケージは非線形回帰の有名なパッケージ mgcv を極値分布に適用したものであり，多くのユーザーにとって扱いやすい．

最後に，回帰分析とは背景が異なるが関連研究として，時間 $t \in [0,1]$ や空間 $s \in \mathbb{R}^2$ に依存する確率過程と極値モデルの融合研究も盛んである．極値確率過程は，Smith (1989)，Einmahl et al. (2016)，de Haan and Zhou (2020) などが知られている．de Haan and Ferreira (2006) にも確率過程の極値理論がまとめられている．空間極値統計学 (spatial extremes) についてはレビュー論文である Davison et al. (2012)，Huser and Wadsworth (2022) に丁寧にまとめられている．Einmahl et al. (2022) は時空間極値モデルを提案し，統計理論の構築にも成功している．

4.2 極値分位点推定

ここでは先行研究が多い GP を考える．3.4 節の設定で再現レベル $z(p)$ は実質的に X の $100(1-p)\%$ 分位点であった．また，閾値 w もデータの最小値と最大値の間にあるのだから X の分位点であると解釈できる．実際に，$P(X > w) \approx k/n$ であるとき，$w \approx z(k/n)$ となる．このとき，再現レベルは

$$z(p) \approx z(k/n) + \frac{\sigma}{\gamma}\left\{\left(\frac{np}{k}\right)^{-\gamma} - 1\right\} \tag{4.1}$$

と表すことができる．確率点 k/n の周りは極値エリアであるもののそれなりにデータ数を確保できる．一方で，確率点 p に相当するエリアにはデータがほとんど存在しない．通常はそのような点の分位点 $z(p)$ は直接推定できないが，(4.1) はそれなりに良い推定が可能な分位点 $z(k/n)$ に極値理論から導かれる定数 $(\sigma/\gamma)\{((np)/k)^{-\gamma} - 1\}$ を足すいわば “外挿 (extrapolation)” の形で構成されていることがわかる．このように外挿を駆使して推定されるような分位点 $z(p)$ は極値分位点 (extreme quantile) と呼ばれ，GEV や GP など極値分布の議論を介さず，直接分位点を推定する文脈で語られることも多い．もちろん極値分位点は再現レベルと本質的に同定義である．

極値分位点は Weissman (1978) を皮切りに様々なモデルの応用がなされている．特に，目的変数 Y の極値分位点を説明変数 X に関連させて回帰モデルの形式での発展がなされ，極値分位点回帰 (extreme quantile regression, EQR) と呼ばれている．その場合は $X = x$ を与えた下での Y の分位点を $z_Y(\cdot|x)$ と書くと，上の結果から直ちに

$$z_Y(p|x) \approx z_Y(k/n|x) + \frac{\sigma(x)}{\gamma(x)}\left\{\left(\frac{np}{k}\right)^{-\gamma(x)} - 1\right\} \tag{4.2}$$

が導かれる．ただし，尺度パラメータと形状パラメータも基本的には説明変数に依存するものとして考えられる．4.1 節でも述べた形状パラメータ $\gamma(x)$ の連続性と符号の問題は当然極値分位点推定でも起きる．そのため，多くの場合で $\gamma(x) > 0$ が仮定される．大まかな方法としては，まず $z_Y(k/n|x)$ を通常の分位点回帰 (Koenker, 2005 等) で推定し，$\gamma(x), \sigma(x)$ を GP 回帰で推定し，(4.2) よりより高位な分位点の推定を行う．最初から $\gamma(x) > 0$ と仮定し，PT を考えるときは条件付き EQR は

$$z_Y(p|x) \approx z_Y(k/n|x)\left(\frac{p}{k/n}\right)^{-\gamma(x)}$$

となる．極値分位点回帰は先行研究が豊富である．Chernozhukov (2005) は外挿は行っていないものの極値統計学の文脈で分位点回帰の裾部分を理論評価している．Daouia et al. (2013)

は EQR のノンパラメトリック推定を確立している．近年では Gnecco et al. (2024)がランダムフォレスト，Allouche et al. (2024)がニューラルネットワーク，Richards and Huser (2024)が深層学習を用いた EQR を開発しており，目覚ましい発展がなされている．

5. クラスターデータに対する極値モデリング

極値統計モデリングは様々な分野で利用されているが，多くの場合，データ構造としてクラスターデータの形式で得られる．代表的なものは気象データで，例えば降雨量データは日本全国で約 1300 地点で観測・集計されている．このようなデータは距離が近いところは関連があり，その情報を取り入れた解析をすることで自然現象の理解を深めること，また予測精度の向上に繋がることもある．単にクラスター間で予測値を平滑化するだけでもわかることは多い．

いま，$X_{ij}, i = 1,\dots,n_j, j = 1,\dots,J$ を J 個のクラスターで i 番目に観測されたデータとする．

クラスターが地理情報を含む場合は，これを空間データととらえることもできる．その場合は，主にコンパクトな空間 $\mathcal{S} \subset \mathbb{R}^2$ における位置 $s \in \mathcal{S}$ に対して，データを $X_i(s)$ と確率過程表現を行うこともある．ただし，降雨量データのように観測地点が固定の場合($\mathcal{S}$ が有限の離散点)で観測時点も等しい場合は単に多変量データ $\{(X_{i1},\dots,X_{iJ}) : i = 1,\dots,n\}$ として扱うことも多い．この場合は，各観測所に付与するダミー変数を説明変数とした場合の回帰分析の文脈でも記述できる．この章ではクラスターデータに対する極値統計モデリングの例を述べる．以降はモデルとして GP を想定するが，GEV でも同様の解析が可能である．

5.1 クラスター統合

ここでは多変量データ $\{(X_{i1},\dots,X_{iJ}) : i = 1,\dots,n\}$ の各 $j = 1,\dots,J$ について，$F_j(x) = P(X_{ij} \le x)$ の裾のモデリングを行う．パラメータ推定にあたり J 個のクラスター情報を統合できれば単純にデータ数が増え，パラメータ推定が安定する．いま，j 番目のクラスターに対して，閾値 w_j としたとき，$X_{1j},\dots,X_{nj}$ の中で $Y_{ij} = X_{ij} - w$ が 0 より大きいものを集め，$\{Y_{1,j},\dots,Y_{k_j,j}\}$ と再ラベリングする．これらのデータが $\mathrm{GP}(\sigma_j,\gamma_j)$ に従うとする：

$$P(Y_{i,j} > y) = \left(1 + \frac{\gamma_j y}{\sigma_j}\right)_+^{-1/\gamma_j}, \quad j = 1,\dots,J.$$

ここで，Yee and Stephenson (2007)などで扱われている代表的なクラスター統合では

$$\gamma = \gamma_1 = \cdots = \gamma_J$$

が仮定される．つまり，形状パラメータは同一であり，尺度パラメータはクラスター固有のものとしている．一般に，σ_j は閾値 w_j に依存するので，σ_j のみの統合は有効な効果が得られない．一方で，形状パラメータは位置尺度不変なので，閾値と尺度パラメータによらず統合を検討できる．また，これは $F_1,\dots,F_J \in \mathcal{D}(G_\gamma)$ という仮定を置いていることになり，J 個のクラスターの元の分布関数そのものは異なるが，極値分布は共通であることを意味する．このような仮定は極値分布論と整合性がとれているため考えやすい．この仮定の下でパラメータ $(\gamma,\sigma_1,\dots,\sigma_J)$ を推定する．その最尤推定量は

$$(\hat\gamma,\hat\sigma_1,\dots,\hat\sigma_J) = \operatorname*{argmax}_{\gamma,\sigma_1,\dots,\sigma_J} \sum_{j=1}^{J}\sum_{i=1}^{k_j} \log f(Y_{i,j}|\gamma,\sigma_j) \tag{5.1}$$

で与えられる．これにより，各 j について

$$P(X_{ij} > y) = P(X_{ij} > w)\left(1 + \frac{\hat{\gamma}(y - w_j)}{\hat{\sigma}_j}\right)^{-1/\hat{\gamma}}$$

から再現レベルを計算できる．ただし，$P(X_{ij} > w_j) \approx k_j/n$ である．

モデルとして GEV を扱う場合は，j 番目のクラスターが持つパラメータを $(\mu_j, \sigma_j, \gamma_j)$ とするとき，GP と同様に γ_j のみクラスターで共通として，(μ_j, σ_j) は統合せず，クラスター固有のパラメータとする．

5.2 統合するクラスターの選定

クラスターの統合は有効な場合はよいが，むやみに統合してよいわけではない．統合するクラスターの選定を何らかの意味で正当化したい．例えば，降雨量データであれば観測地点が近い地域では同じ日に豪雨が観測されるだろう．このように，同じタイミングで極値データが発生するクラスター同士は統合を検討してよさそうである．異なるクラスター間の極値データの同時発生を測る指標として 2.5 節でも用いた裾従属係数を改めて考える．極値データの同時発生を測るための裾従属係数は異なる 2 つのクラスター番号 j, k について，

$$\chi_{jk} = \lim_{q \to 1} P(F_j(X_{ij}) > q | F_k(X_{ik}) > q) = \lim_{q \to 1} \frac{P(F_j(X_{ij}) > q, F_k(X_{ik}) > q)}{1 - q}$$

で定義される．裾従属係数 χ_{jk} が比較的大きい値を持つとき，j 番目と k 番目のクラスターは同時に極値データが発生する確率が高いと判断できる．

いま，主に解析したいクラスター番号を 1 とする．このとき，残りの $J-1$ 個のクラスターのうち，ある $\delta \in (0,1)$ に対して

$$\{j \in \{2, \ldots, J\} | \chi_{1j} > \delta\}$$

となるクラスターのみを集め，改めて $\{(X_{i1}, \ldots, X_{iK}) : i = 1, \ldots, n\}$ と書く．ここで，$K \leq J$ は J 個のうち，1 番目のクラスターとの裾従属係数が高いクラスターの個数である．この K 個のクラスターは統合を検討できるだろう．

本手法について一点補足をしておく．本章では $\{(X_{i1}, \ldots, X_{iJ}) : i = 1, \ldots, n\}$ の多変量確率変数を考えている．したがって，本来は J 変量の多変量分布を検討すべきであろう．実際に，裾従属係数の値が大きいクラスター同士は独立とは考えにくい．さて，多変量分布関数はコピュラ C と $X_{1j}, \ldots, X_{nj}$ の周辺分布関数 F_j を用いて

$$P(X_{i1} \leq x_1, \ldots, X_{iJ} \leq x_J) = C(F_1(x_1), \ldots, F_J(x_J))$$

と表せる．すると対応する密度関数は

$$f(x_1, \ldots, x_J) = c(F_1(x_1), \ldots, F_J(x_J)) \prod_{j=1}^{J} f_j(x_j) \tag{5.2}$$

と表せる．ここで，$c(u_1, \ldots, u_J) = \partial^J C(u_1, \ldots, u_J)/\partial u_1 \cdots \partial u_J$ であり，f_j は F_j の周辺密度関数である．さらに f_j に GP を仮定する場合は

$$f_j(x_j) = \begin{cases} P(X_{ij} < w_j) f_{u,j}(x_j) & x_j < w_j \\ P(X_{ij} \geq w_j) \frac{1}{\sigma}\left(1 + \frac{\gamma(x_j - w_j)}{\sigma}\right)^{-1/\gamma_j - 1} & x_j \geq w_j \end{cases}$$

を考えることになる．ただし，$f_{u,j}$ は $X_{ij} < w_j$ の部分の密度関数であるが，極値解析には用いない形式的なものである．すると，(5.2) の対数尤度は

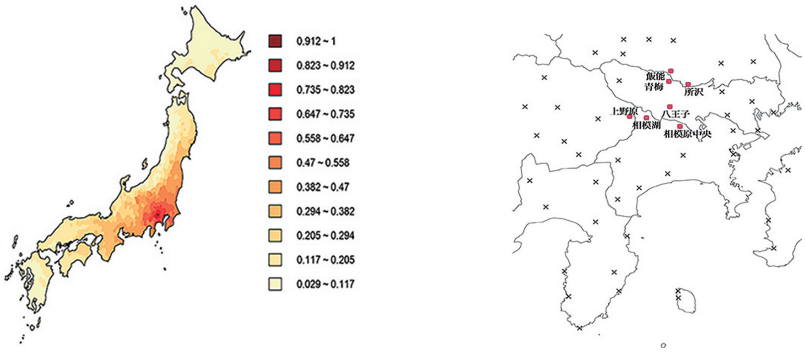

図 6. 左：八王子市と他の 1127 箇所の観測所の裾従属係数．右：八王子市と裾従属係数が 0.75 以上となった観測所（赤点）．× は周辺の観測所．

$$\sum_{i=1}^{n} \log f(X_{i1}, \ldots, X_{iJ}) = \sum_{i=1}^{n} \log c(F_1(X_{i1}), \ldots, F_J(X_{iJ})) + \sum_{i=1}^{n} \sum_{j=1}^{J} \log f_j(X_{ij})$$

となる．従属関数 $c(F_1(X_{i1}), \ldots, F_J(X_{iJ}))$ に含まれる F_j を経験分布関数で置き換えると c の推定と周辺尤度の部分の推定は互いに影響しなくなる．このテクニックは極値解析に関わらず，コピュラを用いた推測ではよく議論されている（Nelsen, 2006）．これより，5.1 節ではクラスター間の従属性を多変量分布としてモデリングせず単に周辺分布の統合を行っているが，これは従属関数 $c(F_1(X_{i1}), \ldots, F_J(X_{iJ}))$ の推定と周辺分布の推定を別で考えていることに相当する．なお，極値解析において従属関数と周辺分布の推定を別で考えることの理論的な正当性は Genest et al.（1995）と Genest and Segers（2009）によって議論されている．

5.3 データ適用

第 2, 3 章と同様の日降雨量データを扱う．全国には約 1300 箇所の雨量観測所があるが，最近開設された場所を除いた $J = 1128$ 箇所の観測所について，東京都八王子市を基準にして全国の各観測所との裾従属係数を調べる．ただし，観測所によって降雨量観測を開始した時点が異なるため，裾従属係数の計算のために用いるデータの観測期間を 2000 年 1 月 1 日〜2023 年 12 月 31 日とした．結果を図 6 に示している．左図から，八王子とその周辺は従属性が高く，離れるほど従属性が低いことがわかる．これは直感的にも明らかで，八王子市で豪雨が降るとき，近隣も雨量は多いが遠方地域では同日に晴れであることもあるだろう．統合クラスターの選定は単純に観測地点間の距離をベースに考えても似たような結果を得るだろうが，裾従属係数はきちんと地理情報を反映した結果を返しているし，株価など地理情報を含まないデータに対しても有効なので汎用性が高い．

八王子との裾従属係数が 0.75 以上であった観測所は埼玉県所沢，飯能，東京都青梅，山梨県上野原，神奈川県相模原中央，相模湖の 6 箇所であった（図 6 右図を参照）．八王子を含めた 7 箇所（$K = 7$）に対して，クラスターの統合解析を行う．それぞれの地域を $j = 1, \ldots, 7$ とし，極値分布として共通の極値分布を持つ，つまり，$F_j \in \mathcal{D}(G_\gamma)$ と仮定する．特にここでは GP を極値モデルとして想定する．まずは各クラスターで閾値を設定する．八王子の閾値は $w = 92.5$ であり，これは上位 79 個のデータを使うことに相当した．八王子以外の観測所も上位 79 個のデータを GP に当てはめる．これに相当する閾値を図 7 に平均残差生存図とともに示している（縦線）．結果からおよそすべての観測所でこの値周辺で比較的安定して線形構造を有している

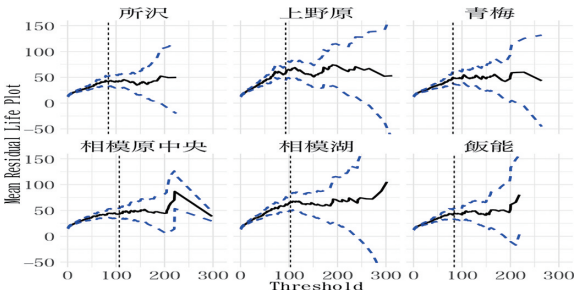

図 7. 飯能，所沢，青梅，上野原，相模原中央，相模湖の降雨量データの平均残差生存図．青線は 95% 信頼区間．

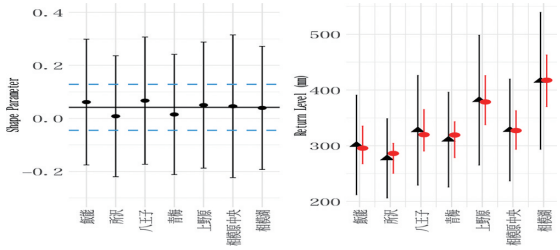

図 8. 左：クラスター統合前後の形状パラメータの推定値(左図)と 50 年再現レベル（右図)．右図では黒が統合前，赤が統合後の結果．

ことがわかった．

図 8(左)は 7 地点毎とクラスター統合後の形状パラメータの推定値である．まず，各地点おける黒点がクラスター統合前の形状パラメータ推定値と 95% 信頼区間(エラーバー)であり，青線がクラスター統合後の推定値(95% 信頼区間)である．クラスター統合後の γ の信頼区間は尤度(5.1)のヘシアンに基づいて構成した．推定値は $\hat{\gamma} = 0.047$ となり，95% 信頼区間は $[-0.04, 0.134]$ となりその幅は 0.17 となった．3.5 節で八王子観測所のみで解析した信頼区間幅は約 0.47 であったため，かなり幅が縮小された(各観測所のエラーバーの長さと破線間の長さを比較しても一目瞭然である)．このように，推定の安定性が向上したと言える．

結果から，所沢，青梅，相模湖は形状パラメータの推定値が若干ではあるが上方修正されていることがわかる．他の観測所は下方修正されているがあまり変化は見られなかった．また，クラスター統合後は信頼区間幅が縮小されたことも見て取れる．これにより，より安定した推測が可能になったと言えるだろう．

図 8(右)には，クラスター統合前(黒)と統合後(赤)の 7 地点での 50 年再現レベル降雨量と 95% 信頼区間を示している．クラスター統合後の再現レベルの信頼区間は(5.1)のヘシアンとデルタ法により構成した．形状パラメータが大きく上方修正された所沢ではクラスター統合後の再現レベルは大きく見積もられた．それ以外の観測所では大きな変化が見られなかったもののクラスター統合により各観測所の予測値が平滑化され，また信頼区間幅も縮小されたことによって，安定した解釈性の高いモデルとなった．

今回は簡単な極値モデリングのクラスター統合を行ったが，重要な留意点がある．例えば，3 つのクラスター(番号を $1, 2, 3$ とする)を考えたとき，$\mathcal{X}_{12}, \mathcal{X}_{23} > \delta$ でも $\mathcal{X}_{13} < \delta$ となる可能性がある．この場合，(X_{i1}, X_{i2}, X_{i3}) を統合できるかどうかは判断が難しい．したがって，この方法のみで統合するクラスターを決定する場合には，基点となるクラスターを 1 つ設定しな

ければならない．ただし，基点を変えると統合できるクラスターが異なる，いわば一意性が担保されないのが問題となる．近年では基点を用いずにクラスター選定とモデリングを同時に行う方法が Rohrbeck and Tawn (2021)，Dupius et al. (2023)によって開発されている．しかし，いずれもクラスター統合の前後でモデルの当てはまりを比較する発見的な手法となっている．そのため，統合後のクラスター数をある程度決定しておく必要があり，クラスター数が大きい場合の解析には向かない．このように，極値データのクラスター統合は重要な研究に位置付けられており，さらなる展開が望まれる手法である．

また，気象データの地域統合に関する研究として，Hosking and Wallis (1993, 1997)に代表される地域頻度解析(Regional Frequency Analysis, RFA)は有名である．RFA では気象学的に均質とみなせる地域のデータをプールする方法であるが，この方法はいわば複数地域の分布関数そのものを同一視する．一方で今回我々が用いた統合モデルは地域間の極値分布は同じであるが分布関数そのものには差異があるというもので，RFA とは若干異なる．これにより，図 8 右図では再現レベルの地域の微小な差異を捉えることができている．ただし，再現レベルがほぼ等しい地域は完全に統合する RFA ベースの解析も検討できるだろう．

また，降雨量データ解析について述べると，事前の解析から標高と降水量は共変関係にあることがわかっている．今回はクラスター以外の情報は考慮していないが，その他の有効な説明変数を用いるとより踏み込んだ解析が期待できるだろう．

6. まとめ

本総説では極値統計学の基礎的な方法論を実データ適用を交えながら議論した．特に，ターゲット変数が 1 次元の場合に焦点を絞り，GEV と GP の 2 つのモデリングを扱った．本総説では触れなかったが，GEV と GP どちらを用いるべきか? と疑問を持つこともあるだろう．一概にどちらが優れていることはないが，Bücher and Zou (2021)はそれぞれのモデルの特徴を踏まえながら比較している．

発展研究としては回帰分析に言及したが，そこでも極値データとして興味ある目的変数はあくまで 1 次元であった．一方で，例えば降雨量データについて，日本では台風時に大きな値を持つことから風速との関連は興味深い．その場合は降雨量，風速が共に大きい値を持つ 2 変量極値データの分析が興味の対象となる．他にも，豪雨時に 2 つの河川が同時に氾濫するリスク，洪水(外水氾濫)と内水氾濫の同時発生(日雨量と時間雨量に起因)や豪雨と高潮の同時発生するリスク，複数の株が同時に極端に高い/低い値を持つリスク予測のモデリングなど従属関係のモデリングも重要である．その場合は多変量極値統計モデル(Multivariate extreme value model)の利用が考えられる．多変量極値分布の先駆けとしては Tiago de Oliveira (1958)，Geffroy (1958/59)，Gumbel (1960)，Sibuya (1960)が挙げられる．また，2 変量正規分布から導かれる極値分布である Hüsler-Reiss model (Hüsler and Reiss, 1989)は最も代表的な多変量極値モデルのひとつである．以降，これまでに多くの多変量極値モデルが提案された (Beirlant et al., 2004)．現在では高次元極値モデルを Engelke and Hitz (2020)がグラフィカルモデリングの枠組みで提案し，Engelke and Ivanovs (2021)で高次元極値モデルのスパース化によるアプローチについて概観している．彼らは次元が膨大な中で極値従属が特徴的な少数の変数を抜き出し，モデルを簡略化する方法をまとめている．特に，条件付き独立を利用した方法は Engelke and Hitz (2020)と深い関連がある．2 変量極値モデルの基本的な性質は北野 (2021)にまとめられている．

本稿では GEV と GP をあくまで独立な確率変数の極値論として紹介した．一方で，この極値統計論は点過程論(Point Process)として記述することもできる (付録 B を参照)．極値統計

に関する点過程については 1 変量の場合は Leadbetter et al. (1983)，Resnick (1987)，Coles (2001)，多変量極値の場合は Resnick (2007)，その訳本の国友・栗栖 (2021) を参照されたい．

最後に，本稿で扱った実データ分析は統計ソフト R の代表的な極値統計分析パッケージ extRemes, ismev, eva を用いた．他にも多くの極値統計のためのパッケージが存在し，Stephenson and Gilleland (2006)，Gilleland et al. (2013)，Belzile et al. (2023) にまとめられている．特に，Belzile et al. (2023) には Python のパッケージについても言及している．和書としては西郷・有本 (2020) が R を用いた極値統計解析について詳しく解説している．

謝　辞

本総説の執筆にあたり，「統計数理」の編集委員，ならびに三名の査読者の方々には数々の貴重なアドバイス，的確なコメントをいただきました．心より御礼申し上げます．本研究は日本学術振興会科学研究費助成事業基盤研究 (B) (課題番号：23K28043)，基盤研究 (C) (課題番号：22K11935) の助成を受けております．

付録 A：GEV と GP の同値性

(2.1) と (3.1) の同値性を述べる．ここでは，(2.1) を仮定した下で (3.1) が成立することのみ示すが，逆も同様である．また，$\gamma \neq 0$ のときのみ証明するが，$\gamma = 0$ のときの証明はより簡単である．

(2.1) より，$P(\max_i X_i \leq x) = P(X_i \leq x)^n = F^n(x)$ について，十分に大きな x, n に対して

$$F^n(x) \approx \exp\left[-\left(1 + \frac{\gamma(x-\mu_n)}{\sigma_n}\right)^{-1/\gamma}\right]$$

であった．ただし，$1 + \gamma(x-\mu_n)/\sigma_n > 0$ となる x を選んでいる．ここで，$\log F^n(x) = n \log F(x) \approx n(1 - F(x))$ より，

$$1 - F(x) \approx -\frac{1}{n}\left(1 + \frac{\gamma(x-\mu_n)}{\sigma_n}\right)^{-1/\gamma}$$

を得る．すると，$\sigma_w = \sigma_n + \gamma(w - \mu_n)$ と置くと以下を得る：

$$\begin{aligned} P(X_i > w + x | X > w) &= \frac{1 - F(w+x)}{1 - F(w)} \\ &\approx \frac{(1 + \gamma(w + x - \mu_n)/\sigma_n)^{-1/\gamma}}{(1 + \gamma(w - \mu_n)/\sigma_n)^{-1/\gamma}} \\ &= (1 + \gamma\frac{x}{\sigma_w})^{-1/\gamma}. \end{aligned}$$

また，別の閾値 $w^* > w$ を用いたとき，$\sigma_{w^*} = \sigma_n + \gamma(w^* - w - \mu_n)$ を用いて

$$P(X_i > w^* + x | X > w^*) \approx \left(1 + \gamma\frac{x}{\sigma_{w^*}}\right)^{-1/\gamma}$$

が成立することが簡単にわかる．つまり，閾値を少し変化させたとき，GP の形状パラメータは不変であるが，尺度パラメータは線形に少しシフトする．

付録 B：GEV の点過程表現

確率変数 $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$ は (2.1) を満たすとする．このとき，十分に大きい x について

$$P\left(\max_i X_i \le x\right) = F^n(x) \approx G_\gamma\left(\frac{x-\mu}{\sigma}\right)$$

となる．このとき，$-\log F^n(x) = -n\log F(x) \approx n(1-F(x))$ より，

$$P(X_i > x) = 1 - F(x) \approx -\frac{1}{n}\log G_\gamma\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{n}\left(1+\frac{\gamma(u-\mu)}{\sigma}\right)^{-1/\gamma}$$

となる．したがって，十分に大きい値 u について

$$p = P(X_i > u) \approx \frac{1}{n}\left(1+\frac{\gamma(u-\mu)}{\sigma}\right)^{-1/\gamma}$$

とすると，$X_1, \ldots, X_n$ の内 u を超過する個数は二項分布 $B(n,p)$ に従う．言い換えると，領域 $A \subset \mathbb{R}$ における点過程 $N_n(A) = |\{i \in \{1,\ldots,n\} : X_i \in A\}|$ と $I_u = [u,\infty)$ について $N_n(I_u) \sim B(n,p)$ が成立する．すると，二項分布がポアソン分布に収束することを利用すると，$n \to \infty$ の下で

$$N_n(I_u) \xrightarrow{D} N(I_u) \sim Po(\lambda)$$

を得る．ただし，

$$\lambda = \left(1+\frac{\gamma(u-\mu)}{\sigma}\right)^{-1/\gamma}$$

である．ここで，$P(\max_i X_i < x) = P(N_n(I_x) = 0)$ なので，確かに

$$P\left(\max_i X_i < x\right) = P(N_n(I_x)=0) \to P(N(I_x)=0) = \exp[-\lambda] = G_\gamma\left(\frac{x-\mu}{\sigma}\right)$$

を満たす．これが GEV のポアソン点過程表現である．同様に，GP も点過程表現できる（Coles, 2001）．

参 考 文 献

Allouche, M., Girard, S. and Gobet, E. (2024). Estimation of extreme quantiles from heavy-tailed distributions with neural networks, *Statistics and Computing*, **34**, https://doi.org/10.1007/s11222-023-10331-2.

Barder, B. and Yan, J. (2020). eva: Extreme Value Analysis with Goodness-of-Fit Testing, R package version 0.2.6., https://CRAN.R-project.org/package=eva（最終アクセス日 2025 年 6 月 4 日）.

Beirlant, J. and Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions, *Journal of Multivariate Analysis*, **89**, 97–118.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004), *Statistics of Extremes*: *Theory and Applications*, John Wiley & Sons, Chichester.

Belzile, L., Dutang, C., Northrop, P. and Opitz, T. (2023). A modeler's guide to extreme value software, *Extremes*, **26**, 595–638.

Bücher, A. and Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution, *Extremes*, **20**, 839–872.

Bücher, A. and Zhou, C. (2021). A horse race between the block maxima method and the peak-over-threshold approach, *Statistical Science*, **36**, 360–378.

Caeiro, F. and Gomes, M. I. (2011). Asymptotic comparison at optimal levels of reduced-bias extreme value index estimators, *Statistica Neerlandica*, **65**, 462–488.

Caires, S. (2011a). Extreme value analysis: Wave data, JCOMM Technical Report, **57**, World Meteorological Organization, Geneva.

Caires, S. (2011b). Extreme value analysis: Still water level, JCOMM Technical Report, **58**, World Meteological Organization, Geneva.

Castro-Camilo, D., Huser, R. and Rue, H. (2022). Practical strategies for generalized extreme value models for extremes, *Environmetrics*, **33**, https://doi.org/10.1002/env.2742.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modeling of sample extremes, *Journal of the Royal Statistical Society Series C*, **54**, 207–222.

Chernozhukov, V. (2005). Extremal quantile regression, *Annals of Statistics*, **33**, 806–839.

Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, New York.

Coles, S. G., Heffernan, J. and Tawn, J. (1999). Dependence measures for extreme value analyses, *Extremes*, **2**, 339–365.

Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, **19**, 2557–2589.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds, *Journal of the Royal Statistical Society Series B*, **52**, 393–442.

Davison, A. C., Padoan, S. and Ribatet, M. (2012). Statistical modelling of spatial extremes (with discussion), *Statistical Science*, **27**, 161–186

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory — An Introduction*, Springer, New York.

de Haan, L. and Zhou, C. (2020). Trends in extreme value indices, *Journal of the American Statistical Association*, **116**, 1265–1279.

de Haan, L. and Zhou, C. (2024). Bootstrapping extreme value estimators, *Journal of American Statistical Association*, **119**, 382–393.

Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Annals of Applied Probability*, **14**, 1179–1201.

Dupuis, D. J., Engelke, S. and Trapin, L. (2023). Modeling panels of extremes, *Annals of Applied Statistics*, **17**, 498–517.

Einmahl, J. H., de Haan, L. and Zhou, C. (2016). Statistics of heteroscedastic extremes, *Journal of the Royal Statistical Society Series B*, **78**, 31–51.

Einmahl, J. H., Ferreira, A., de Haan, L., Neves, C. and Zhou, C. (2022). Spatial dependence and space-time trend in extreme events, *Annals of Statistics*, **50**, 30–52.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes, *Journal of the Royal Statistical Society Series B*, **82**, 871–932.

Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes, *Annual Review of Statistics and Its Application*, **8**, 241–270.

Fisher, R. A. and Tippet, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest members of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180–190.

Geffroy, J. (1958/59). Contribution à la théorie des valeurs extrêmes, *Publications de l'Institut de Statistique de l'Université de Paris*, **7**, 37–121, and **8**, 123–184.

Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas, *Annals of Statistics*, **37**, 2990–3022.

Genest, C., Ghoudi, K. and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82**, 543–552.

Gilleland, E. and Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R, *Journal of Statistical Software*, **72**, 1–39.

Gilleland, E., Ribatet, M. and Stephenson, A. (2013). A software review for extreme value analysis, *Extremes*, **16**, 103–119.

Gnecco, N., Terefe, EM. and Engelke, S. (2024). Extremal random forests, *Journal of American Statistical Association*, **119**, 3059–3072.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum of d'unesérie Aléatorie, *Annals of Mathematics*, **44**, 423–453.

Goegebeur, Y., Guillou, A. and Stupfler, G. (2015). Uniform asymptotic properties of a nonparametric regression estimator of conditional tails, *Annales de I'Institut Henri Poincaré-Probabilitésés et Statistiques*, **51**, 1190–1213.

Gomes, M. I. and Martins, M. J. (2002). "Asymptotically unbiased" estimators of the extreme value index based on external estimation of the second order parameter, *Extremes*, **5**, 5–31.

Gomes, M. I. and Pestana, D. D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator, *Journal of American Statistical Association*, **102**, 280–292.

Guedes Soares, C. and Scotto, M. G. (2004). Application of the r largest-order statistics for long-term predictions of significant wave height, *Coastal Engineering*, **51**, 387–394.

Gumbel, E. J. (1960). Multivariate extremal distributions, *Bulletin of International Statistical Institute*, **39**, 471–475.

Haigh, I. D., Nicholls, R. and Wells, N. (2010). A comparison of the main methods for estimating probabilities of extreme still water levels, *Coastal Engineering*, **57**, 838–849.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **13**, 331–341.

Hosking, J. R. M. and Wallis, J. R. (1993). Some statistics useful in regional frequency analysis, *Water Resources Research*, **29**, 271–281.

Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis*: *An Approach Based on L-Moments*, Cambridge University Press, Cambridge.

Huser, R. and Wadsworth, J. L. (2022). Advances in statistical modeling of spatial extremes, *WIREs Computational Statistics*, **14**, https://doi.org/10.1002/wics.1537.

Hüsler, J. and Reiss, R. D. (1989). Maxima of normal random vectors: Between independence and complete dependence, *Statistics and Probability Letters*, **7**, 283–286.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quarterly Journal of the Royal Meteorological Society*, **81**, 158–171.

北野利一 (2020). 再現期間再探訪, 極値理論の工学への応用 (18), 統計数理研究所共同研究リポート, No.445, 63–70.

北野利一 (2021). 成分毎の最大値と閾値を超過する多変量極値：それらの相互関係, それらの単純極値分布と乱数生成法, 日本統計学会誌, **51**, 123–156.

北野利一, 加藤紗也, 平松健太郎 (2025). 年最大日降水量とピーク流量の従属性と治水計画におけるカバー率, 土木学会論文集, **81**, https://doi.org/10.2208/jscejj.24-16135.

Koenker, R. (2005). *Quantile Regression*, Cambridge University Press, Cambridge.

国友直人, 栗栖大輔（訳）(2021).『極値現象の統計分析：裾の重い分布のモデリング』, 朝倉書店, 東京.

Leadbetter, M. R., Lindgren, G. and Rootzn, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Lin, R., Leng, C. and You, J. (2022). Semiparametric tail index regression, *Journal of Business & Economic Statistics*, **40**, 82–95.

Meeker, W. Q., Escobar, L. A. and Pascual, F. G. (2021). *Statistical Methods for Reliability Data*, 2nd ed., Wiley, New Jersey.

Murphy, C., Tawn, J. A. and Varty, Z. (2024). Automated threshold selection and addoxiated inference uncertainty for univariate extremes, *Technometrics*, **67**, 215–224.

Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer, New York.

Northrop, P. J., Attalides, N. and Jonathan, P. (2017). Cross-validatory extreme value threshold selec-

tion and uncertainty with application to ocean storm severity, *Journal of the Royal Statistical Society Series C*: *Applied Statistics*, **66**, 93–120.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*, Springer, New York.

Resnick, S. I. (2007). *Heavy-tail Phenomena*: *Probabilistic and Statistical Modeling*, Springer, New York.

Richards, J. and Huser, R. (2024). Extreme quantile regression with deep learning, arXiv, https://doi.org/10.48550/arXiv.2404.09154.

Rohrbeck, C. and Tawn, J. A. (2021). Bayesian spatial clustering of extremal behavior for hydrological variables, *Journal of Computational and Graphical Statistics*, **30**, 91–105.

西郷達彦, 有本彰雄 (2020).『R による極値統計学』, オーム社, 東京.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *REVSTAT - Statistical Journal*, **10**, 33–60.

Sibuya, M. (1960). Bivariate extreme statistics, I, *Annals of Institute of Statistical Mathematics*, **11**, 195–210.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.

Smith, R. L. (1986). Extreme value theory based on the r largest annual events, *Journal of Hydrology*, **86**, 27–43.

Smith, R. L. (1987). Estimating tails of probability distributions, *The Annals of Statistics*, **15**, 1174–1207.

Smith, R. L. (1989). Extremevalue analysis of environmental time series: An application to trend detection in ground-level ozone, *Statistical Science*, **4**, 367–377.

Spearing, J. Tawn, J. Irons, D. and Paulden, T. (2023). A framework for statistical modelling of the extremes of longitudinal data, applied to elite swimming, arXiv, https://doi.org/10.48550/arXiv.2306.12419.

Stephenson, A. G. and Gilleland, E. (2006). Software for the analysis of extreme events: The current state and future directions, *Extremes*, **8**, 87–109.

Stephenson, R. A. (2018). ismev: An Introduction to Statistical Modeling of Extreme Values, R package version 1.42, https://CRAN.R-project.org/package=ismev（最終アクセス日 2025 年 6 月 4 日）.

高橋倫也, 志村隆彰 (2016).『極値統計学(ISM シリーズ：進化する統計数理)』, 近代科学社, 東京.

Tawn, J. A. (1988). An extreme-value theory model for dependent observations, *Jornal of Hydrology*, **101**, 227–250.

Tiago de Oliveira, J. (1958). Extremal distributions, *Faculdade de Ciencias de Lisboa. Série A*: *Matemática*, **7**, 219–227.

von Mises, R. (1936). La distribution de la plus grande de n valeurs, *Revue de Mathématiques de l'Union Interbalkanique*, **1**, 141–160.

Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection, *Technometrics*, **58**, 116–126.

Wahl, T., Haigh, I. D., Jensen, J. and Pattiaratchi, C. (2013). Estimating extreme water level probabilities: A comparison of the direct methods and recommendations for best practise, *Coastal Engineering*, **81**, 51–66.

Wang, H. and Tsai, C. L. (2009). Tail index regression, *Journal of the American Statistical Association*, **104**, 1233–1240.

Weibull, W. (1939). A statistical theory of the strength of materials, *Ingeniörsvetenskapsakademiens Handlingar* (*Royal Swedish Academy of Engineering Sciences*), **151**, 1–45.

Weibull, W. (1951). A statistical distribution function of wide applicability, *Journal of Applied Me-*

chanics, **18**, 293–297.

Weissman, I. (1978). Estimation of parameters and larger quantiles based on the k largest observations, *Journal of American Statistical Association*, **73**, 812–815.

Yee, T. W. and Stephenson, A. G. (2007). Vectorgeneralized linear and additive extreme value models, *Extremes*, **10**, 1–19.

Youngman, B. D. (2019). Generalized additive models for exceedances of high threshilds with an application to return level estimation for U. S. wing gusts, *Journal of the American Statistical Association*, **114**, 1865–1879.

Youngman, B. D. (2022). evgam: An R, package for generalized additive extreme value models, *Journal of Statistical Software*, **103**, 1–26.

Zhong, P., Huser, R. and Opitz, T. (2022). Modeling nonstationary temperature maxima based on extremal dependence changing with event magnitude, *Annals of Applied Statistics*, **16**, 272–299.

316 Proceedings of the Institute of Statistical Mathematics Vol. 73, No. 2, 287–316 (2025)

Statistical Modeling Using Extreme Value Theory

Takuma Yoshida[1] and Toshikazu Kitano[2]

[1]Graduate School of Science and Engineering, Kagoshima University
[2]Department of Architecture, Nagoya Institute of Technology

In applied fields dealing with various real-world phenomena such as natural disasters caused by heavy rainfall or earthquakes, financial risks in finance, and product lifespans, estimating the probabilities of extremely large or small values within the overall data is a critical issue from the perspective of reliability assessment and risk management. The statistical topic that addresses this issue is the prediction of the maximum or minimum values of the data, or quantiles close to these extremes. For this purpose, it is essential to construct probabilistic models that focus solely on the tail behavior. Extreme value statistics provide the methodological framework for such analyses, offering a mathematically rigorous approach to modeling the tails of distributions where rare events occur. Naturally, however, modeling for real-world data requires tailored approaches that consider the unique characteristics of the data. This paper provides an overview of the fundamental concepts, statistical theories, and modeling techniques in extreme value statistics. Special attention is given to the concept of return periods, which play a critical role in understanding the frequency characteristics of phenomena based on the analysis results derived from applying extreme value statistical methods to real data. In addition, data sets where extreme value statistics are applied, such as meteorological and financial market data, are often obtained in the form of clustered data. Therefore, this paper also discusses methodological approaches to extreme value statistical modeling for clustered data. Finally, we illustrate the application of extreme value statistical modeling to rainfall data and present relevant R packages to demonstrate the practical implementation.

Key words: Extreme value theory, clustered data, generalized extreme value distribution, generalized pareto distribution, return level, return period.