2006/06/01
岐阜大学　本城勇介
複合技術研究所　篠田昌弘

付録 2　統計的手法による作用モデルの構築

要旨　地震動，風，雨，雪，波，温度等による土木構造物への設計作用を決定しようとすると、そこで問題になるのは日常的に生起している作用ではなく、何十年、何百年、 場合によっては数千年に一度生起するような事象を対象にする必要がある。 しかし一方で我々が入手できる信頼できる観測データは、普通数十年、 長くても数百年である。我々はこのようなデータから外挿によって 求めたい極値を推定する。このような問題は極値統計学で扱われ、 従来 1920 年代の Fisher と Tippett による極値分布の発見にはじまり、 1960 年代に Gumbel により集大成された応用に関する理論がその主流を 占めてきた。しかし 1970 年代に入り、データのうち、ある閾値より 大きいデータのみを取り出して解析する POT 解析 (Peaks over threshold analysis) が Balkeman、de Haan や Pickands らによって理論的に確立され、 極値統計量の解析は、新しい時代に入った。 この方法は近年、金融工学や保険の分野でリスクの計量に広く用いられ始め、 user friendly なソフトウエアの開発もあいまって、飛躍的に発展している。 ここに示すのは、このような POT 解析に至るまでのごく簡単な極値統計学 のまとめと、これを説明するために作成した幾つかの簡単な解析例である。

1. 概論

以下は「日本統計学会会報」（1999 年 12 月号）の『統計学の現状と今後』欄に掲載された極値統計学者、高橋倫也(1999)の「極値統計学へのお誘い」と言う文章からの引用である。極値統計学の概要を歴史的な発展に即して簡潔に記述しており、きわめてすぐれた文書であるので、ここに本論の「概論」として挙げる。（途中省略下部分と、一部変更した箇所がある。）

「おまえは何を研究している？」と聞かれたらどう答えようかと考えたことがある．状況に応じて相手が望むであろうことを答えればいいのであるが，それでは面白くない．そこで「しっぽ(tail)の研究を！」というのはどうであろうか？しかし，私がこのようなことを言うと「とかげの？」と言われそうである．最近，ある人から「統計学者には middle man と tail man がいる．」と言った人がいることを教えていただいた．

通常の統計学では母集団分布の中心が主な推測の対象になるが，極値統計学ではそれは分布の裾(tail)である．極値統計学では，データの全部を利用することは少なくデータの上位（または下位）部分のみを利用することがほとんどで，通常の統計学と違い正規分布は（推定量の漸近分布として以外は）出てこない．また，裾の重たい(heavy tail)分布を扱うことが多い．

極値統計学で扱う典型的な問題は，次の様なものである：非常に強固な防波堤を作りたい．そのために，利用可能な過去百年間の潮位観測データから，今後１万年間の最高潮位を予測せよ．また，６ｍの防波堤を作ったとしたときそれを越える高潮の起きる確率を推定せよ．それらの結果から防波堤の高さを決定しよう．

この問題は，与えられた（一部の）データを利用して全体またはある範囲のデータの最大値を予測せよ，というものである．（確率の推定も同様に扱えるので以下省略する．）この種の問題は工学の多くの分野にある．例えば，水文学では１００年最大降水量を，腐食工学では器機全体での最大腐食を，建築工学では５０年最大風速や地震の最大震度を，また信頼性工学では最大ストレスを，それぞれ与えられた（部分的な）観測データから予測しなければならない．従って，工学の分野では古くから極値統計学のユーザは多い．一方，保険数学の分野では以前から非常に大きな賠償金の予測が問題になっていた．

これ等の問題を統計学の言葉で表せば次のようになる：未知の母集団分布からのデータを用いて，その母集団分布の１に非常に近い確率に対する確率点(quantile)を推定せよ．このような確率点を推定するためにはいわゆるデータの外挿を行わなければならず，困難な問題になる．この問題を解くために，極値統計学では「未知の母集団分布はある極値分布の吸引領域に属す」という仮定を置く．

日本でこの分野を専攻する統計学者は少なく，また世界でもそれほど多くはない．そこで，極値統計学関係のワークショップに参加すると，この分野の指導的な（いわゆる，有名な）研究者のほとんどに会うことが出来る．

極値統計学の過去： 極値統計学の「現状と今後」を考える上で，この分野の統計理論（極値理論）の発展の歴史をながめ今後について考えてみたい．この分野も応用（ユーザー，特に水文学者）と理論（統計学者）の相互作用で発展してきた．以下ではこの分野の統計理論の発展を主に述べる．

先ず，独立同一分布の確率変数の最大値（極値統計量）の漸近分布（極値分布）が，Fisher & Tippett(1928)により求められた．彼らは極値分布には，現在（逆）ワイブル分布，グンベル分布，フレッシェ分布と呼ばれている３つの型があることを見つけた．また，正規分布からの極値統計量の分布の極値分布への収束が遅いことも見いだした．von Mises(1936)は，分布関数が極値分布の吸引領域に属すための十分条件を与えた．彼の条件から，統計学の教科書でお馴染みのほとんどの連続分布が極値分布の吸引領域に属すことが分かる．例え

ば，一様分布，ベータ分布等が（逆）ワイブル分布の，正規分布，ワイブル分布，ガンマ分布，対数正規分布等がグンベル分布の，そしてｔ分布，パレート分布等がフレッシェ分布の吸引領域に属す．これらの研究を数学的に完成させたのは Gnedenko(1943)である．彼は安定分布と極値分布の関連から正則変動関数の理論等を用いて，分布関数が極値分布の吸引領域に属すための必要十分条件を与えた．

極値理論の応用（特に工学）に関する研究を大成したのは Gumbel(1958)である．彼の方法の特徴は，年最大値と言われるある期間またはある領域等での最大値データ（極値データ）を用いる事にある．得られた極値データに３つの型の極値分布のどれかを当てはめ，パラメータの推測を行う．また，データの診断には極値確率紙等を用いる事を多くの例と図で示している．一方，信頼性理論では Weibull(1951)による最小データに関する有名な研究がある．

極値データがどの型の極値分布に適合するのかをあらかじめ決めるのは難しい．そこで，３つの型の極値分布を一つの式（von Mises-Jenkinson 表現）で表した一般極値分布を利用するデータ解析法の研究が，１９５０年代にイギリスを中心にされるようになった．

極値データのみをデータから選んで利用するのは，他のデータの持っている情報を捨てる事になり情報の損失が甚だしい．そこで，極値データではなくある閾値以上のデータを使う手法が水文学で研究されていた．これは POT(peaks over threshold)解析と呼ばれ，経験上ある閾値以上のデータの分布が指数分布で近似出来ることが知られていた．この理論的な裏付けをしたのが，Balkeman & de Haan(1974)と Pickands(1975)による一般パレート分布の発見である．すなわち，分布関数が一般極値分布（（逆）ワイブル分布，グンベル分布，フレッシェ分布）の吸引領域に属すことは，ある閾値以上のデータの従う分布が一般パレート分布（ベータ分布（の一部），指数分布，パレート分布）で近似できることと同値である．以後，一般パレート分布による極値データ解析が極値統計学の主流になってきた．

極値統計学の現状： 上で述べた様に理論が先行しがちであったが，最近では次の様な極値データ解析の研究が盛んに行われている．

１変量極値データ解析：未知の母集団分布がある一般極値分布の吸引領域に属す（すなわち，分布の裾は一般パレート分布で近似できる）と仮定する．この一般極値分布（一般パレート分布）の形状パラメータの逆数を分布の tail-index と言う．特に，tail-index の値が０の場合がグンベル分布（指数分布）の場合になる．精度の良い確率点を推定するために，tail-index の推定に関する研究が行われている．

ある施設の周囲での風速の分布が問題である場合を考えよう．風速の分布の裾がベータ分布で近似されるならば安全，指数分布で近似されるならば要注意，パレート分布で近似されるならば危険と解釈できる．従って，この場合は tail-index が０であるかどうかの検定方法の研究が重要である（Teugels，1999).

環境統計学で汚染物質の集中を問題にするとき，極値理論を用いる必要がある．そこでは時系列データや共変量のあるデータを解析しなければならない．この研究は Smith(1989)等によりされている．また，時系列がらみで最近流行の金融工学関連の研究もある． これらの研究では一般パレート分布が用いられる．このとき，最適な閾値の決定は重要で難しい問題になる．これは，まるで「蛇のどこまでがしっぽでどこからが胴体か？」と問われているような気分に私をさせる問題である．

ところで，極値分布への収束は一般に遅い，しかし安定分布への収束も遅い(de Haan, 1999)と言う指摘は興味深い．

現在，極値理論の研究はアメリカやヨーロッパ各国の有力な研究者を中心にされている．それぞれのグループは，特色のある応用分野（水文学，環境保全，自然災害，異常気象，信頼性工学，保険数学，金融工学等）を視野に入れた研究を精力的に行なっている．

極値統計学の今後： 極値理論の今後について，この分野の多くの研究者は次の様に考えていると思われる．複雑でかつ不確実で予測しがたくなっていくこの世界で，種々のリスク

を予測・評価することが必要な場面・局面が増えてくるだろう．そのとき，この理論はリスクの予測・評価の統計学的（科学的）な手法を提供する有力な理論になるだろう． 極値理論の現状と今後について書くつもりであったが，筆者の力不足と会員の多くの方がこの分野に馴染みが薄いと考え，主に（私の考えている）極値理論の過去と現状についての報告になり，今後についてはほとんど述べられなかった．内容が「竜頭蛇尾」になったことをお詫びしなければならない． この分野に興味を持たれた方が，研究に参加し今後について思いを巡らせていただけたら幸いである．

なお、極値統計学の作用データの整理について解説した邦文の文献としては、日本建築学会(2004)「建築物加重指針・同解説第 4 版」、土木学会(2001)「新しい波浪算定法とこれからの海域施設の設計法」pp.65-84、水文・水資源学会(2003)「水文・水資源ハンドブック」pp.228-248 などが参考になる。

2. 極値分布モデル

2.1 モデルの構成

確率変数 X_i ($i = 1,\ldots, n$)がそれぞれ独立同一分布(*iid*: independently identically distributed) F に従うとき，次のような確率変数 M_n を最大値(maxima)と定義する．

$$M_n = \max\{X_1,\ldots, X_n\} \tag{2.1.1}$$

一般的には，X_i は規則的な時間間隔で計測された代表値であり，日気温や 1 時間降水量などを指す．M_n は観測数 n における最大値を指し，最大日気温や最大 1 時間降水量などを指す．観測数 n が 1 年間の観測数と一致する場合には，得られる最大値は年最大値となる．M_n の確率分布関数は，次のように表される．

$$\begin{aligned}\Pr\{M_n \le x\} &= \Pr\{X_1 \le x,\ldots, X_n \le x\} \\ &= \Pr\{X_1 \le x\}\times\cdots\times\Pr\{X_n \le x\} \\ &= \{F(x)\}^n\end{aligned} \tag{2.1.2}$$

極値分布モデルは、上記のように、独立同一分布(*i.i.d.*)に従い、ある期間に発生する確率変数の最大値の分布を研究することを基本としている。この基本的な枠組みを十分に理解し、その限界を踏まえてこの方法を用いることが重要である。なお、この節の説明は、Coles(2001)を主に参考にしている。

2.2 極値分布

定理 2.1　$\{a_n > 0\}$，$\{b_n\}$の数列が存在し，$n \to \infty$の条件下では，

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le x\right\} \to G(x)$$

が成り立つ．ここで，G は退化しない分布関数である．G は以下の分布のいずれかに属する．

$$\text{I}: G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \qquad -\infty < x < \infty$$

$$\text{II}: G(x) = \begin{cases} 0, & z \le b \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases}$$

$$\text{III}: G(x) = \begin{cases} \exp\left[-\left\{-\left(\frac{x-b}{a}\right)^{\alpha}\right\}\right], & z < b \\ 1 & z \ge b \end{cases}$$

ここで，$a > 0$，b，αはパラメータである．

上記の定理は，変換したサンプルの最大値$(M_n - b_n)/a_n$ は極値分布 I，II，III のいずれかに収束することを示しており，それぞれ Gumbel，Fréchet，Weibull 分布と呼ばれる．それぞれの分布には位置母数 a と尺度母数 b を有し，Fréchet と Weibull 分布は形状母数αを有する．以上の結果は、1928 年に Fisher と Tippett により発表され。これが極値統計学確立の端緒となった。

2.3 一般化極値分布

ここでは，定理 1 により示された 3 つの極値分布を統一的に記述する確率分布を示す．位置母数をμ，尺度母数をσ，形状母数をξとし，$1+\xi(x-\mu)/\sigma>0$ の下では，Gumbel，Fréchet，Weibull 分布は，以下に示す一般化極値分布として表現できる．

$$G(x)=\exp\left[-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\right],\qquad -\infty<x<\infty \tag{2.3.1}$$

ここで，$-\infty<\mu<\infty$，$\sigma>0$，$-\infty<\xi<\infty$である．Fréchet 分布は$\xi>0$，Weibull 分布は$\xi<0$ として表現できる．$\xi=0$ の場合は$\lim_{\xi\to 0}(1+\xi x)^{-1/\xi}=e^{-x}$ であることから、以下に示す Gumbel 分布と一致する．

$$G(x)=\exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\},\qquad -\infty<x<\infty \tag{2.3.2}$$

一般化極値分布を用いると，**定理 2.1** は以下のように書き改めることができる．

定理 2.2　$\{a_n>0\}$，$\{b_n\}$の数列が存在し，$n\to\infty$の条件下では，

$$\Pr\left\{\frac{M_n-b_n}{a_n}\le x\right\}\to G(x)$$

が成り立つ．ここで，G は退化しない分布関数である．G は以下の分布に属する．

$$G(x)=\exp\left[-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\right],\qquad -\infty<x<\infty$$

ここで，$1+\xi(x-\mu)/\sigma>0$，$-\infty<\mu<\infty$，$\sigma>0$，$-\infty<\xi<\infty$である．

この極値分布の一般化表現は、1955 年に Jenkinson により得られたもので、Jenkinson 表現と呼ばれることがある。この表現では、三つの分布を一つの形式で表現できるので、パラメータ推定のとき便利である。

ある確率 F に関する累積分布関数の逆関数の値を、クオンタイル値、またはクオンタイルと言う。一般化極値分布のクオンタイルは，$\xi\neq 0$ では，次式により求めることができる．

$$x_p=\mu-\frac{\sigma}{\xi}\left[1-\{-\ln(F)\}^{-\xi}\right] \tag{2.3.3}$$

また，$\xi=0$ では，次式により求めることができる．

$$x_p=\mu-\sigma\ln[-\ln(F)] \tag{2.3.4}$$

2.5 極値分布モデリング

2.5.1 再現レベル

一般化極値分布における N 年確率量は，以下のように求めることができる．

$$z_N = \mu - \frac{\sigma}{\xi}\left[1 - \left\{-\ln\left(1 - \frac{1}{N}\right)\right\}^{-\xi}\right] \quad (\xi \neq 0) \tag{2.5.1.1}$$

$$z_N = \mu - \sigma \ln\left[-\ln\left(1 - \frac{1}{N}\right)\right] \quad (\xi = 0) \tag{2.5.1.2}$$

上記の関係を用いて、ある作用の N 年超過確率値を求めることができる。実際 100 年超過確率等は、作用の一つの指標としてよく用いられる。

2.5.2 モデルの妥当性の確認

モデルの妥当性を確認するためには，Probability plot や Quantile plot などがある．Probability plot や Quantile plot が線形であれば推定したモデルは観測データによく当てはまっていると言える．閾値 u に対する超過変数を $y_{(1)} \leq \cdots \leq y_{(N)}$とし，推定モデルを $\hat{H}$ とすると，Probability plot は以下の組み合わせで作成することができる．

$$\left\{\left(\frac{i}{N+1}, \hat{H}\left(y_{(i)}\right)\right); i = 1, \ldots, N\right\} \tag{2.5.2.1}$$

ここで，推定モデル $\hat{H}$ は以下のように求めることができる．

$$\hat{H}(y) = \exp\left[-\left\{1 + \hat{\xi}\left(\frac{y - \hat{\mu}}{\hat{\sigma}}\right)\right\}^{-\frac{1}{\hat{\xi}}}\right] \tag{2.5.2.2}$$

また，Quantile plot は以下の組み合わせで作成することができる．

$$\left\{\left(\hat{H}^{-1}\left(\frac{i}{N+1}\right), y_{(i)}\right); i = 1, \ldots, N\right\} \tag{2.5.2.3}$$

ここで，$\hat{H}^{-1}$ は以下のように求めることができる．

$$\hat{H}^{-1}\left(\frac{i}{N+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \left\{-\ln\left(\frac{i}{N+1}\right)\right\}^{-\hat{\xi}}\right] \tag{2.5.2.4}$$

なおこのとき、与えられた個々のデータを、どのようなクオンタイル値に対応させるかと言うことについて、いくつかの考え方がある。この問題は、plotting position の問題と言われ、代表的なものに、Thomas plot と Hazen plot がある。それぞれは、次の通りである。

Thomas plot: $$y_{(i)} = \hat{H}^{-1}\left(\frac{i}{N+1}\right) \tag{2.5.2.5}$$

Hazen plot: $$y_{(i)} = \hat{H}^{-1}\left(\frac{i - 0.5}{N}\right) \tag{2.5.2.6}$$

最近は、Thomas plot がよく用いられるようである。

3. 閾値モデル

3.1 閾値 u の超過分布

第 1 節でもふれたが、多くの計測データが存在するとき、そのほとんどを捨てて、ある期間やブロックの最大値だけを取り出す極値統計解析の方法は、データに関しては著しく不経済である。全データから大きいものをいくつか選び出し（閾値以上のデータ)、それに基づいた解析をする方が合理的に思われる。このような事情から発達してきた極値統計解析の方法が、閾値モデルである。

確率変数 X_i ($i = 1,\ldots, n$)がそれぞれ独立同一分布(*iid*: independently identically distributed) F に従うとき，ある閾値 u を超える任意の確率変数 X_i の条件付き確率は $y > 0$ という条件の下で以下のように表される．

$$\begin{aligned}\Pr\{X > u + y | X > u\} &= \frac{\Pr\{X > u + y, X > u\}}{\Pr\{X > u\}} \\ &= \frac{\Pr\{X > u + y\}}{\Pr\{X > u\}} \\ &= \frac{1 - F(u + y)}{1 - F(u)}\end{aligned} \tag{3.1.1}$$

式(3.1.1)は，確率分布 F が与えられれば，閾値 u の超過分布が得られることを示している．

3.2 一般化パレート分布

定理 3.1　確率変数 X_i ($i = 1,\ldots, n$)がそれぞれ独立同一分布 F に従い，次のような確率変数 M_n を最大値(maxima)と定義する．

$$M_n = \max\{X_1,\ldots, X_n\} \tag{3.2.1}$$

独立同一分布 F が**定理 2.2** を満たし，且つ n が大きい場合には，

$$\Pr\{M_n \le z\} \approx G(z) \tag{3.2.2}$$

ここで，G はあるパラメータμ，$\sigma > 0$，ξ の下で以下のように表すことができる．

$$G(x) = \exp\left[-\left\{1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\right] \tag{3.2.3}$$

n が十分大きいとき，$X > u$ における $Y = X - u$ の条件付き確率分布は近似的に以下のように表すことができる．

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{3.2.4}$$

ここで，$y > 0$，$\left(1 + \frac{\xi y}{\tilde{\sigma}}\right) > 0$ であり，

$$\tilde{\sigma} = \sigma + \xi(u - \mu) \tag{3.2.5}$$

である．

式(3.2.4)で表される確率分布は一般化パレート分布という．**定理 3.1** は最大値分布が一般化極値分布で近似できるとき，閾値超過分布は対応した一般化パレート分布を有することを示している．さらに，閾値を超過した一般化パレート分布の母数は，関連した一般化極値分布によって一意に決定できる．特に，式(3.2.4)の形状母数ξは式(3.2.3)の形状母数ξと一致する．
式(3.2.4)において，$\xi = 0$ の場合には，

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \qquad y > 0 \tag{3.2.6}$$

となり，指数分布と一致する．
ある閾値以上のデータが、指数分布に従うことは水文学者の間では、以前から経験的に知られていた。しかし上記の定理を厳密に証明したのは、Balkema and de Haan(1974)と Pickands(1975)である。この結果、極値統計解析は新しい段階に入った。

3.3 閾値超過モデリング

3.3.1 閾値（超過数）の設定

閾値（超過数）の設定にあたっては，次のような点が指摘されている．

- 閾値が大きすぎる（超過数が小さすぎる）と異なる閾値（超過数）に対して形状母数が激しく変動し，不安定である．
- 中間的な閾値（超過数）に対しては形状母数は安定し，真の値に近い値を与える．
- 閾値が小さすぎる（超過数が大きすぎる）と超過分布の一般化パレート分布に関する収束の前提が成り立たなくなるため，真の値から離れた値を与える．

限られた観測データを用いて適切な近似を行うためには、上記項目を考慮して，閾値をできるだけ小さく設定することも重要となる．一般的に用いられている閾値の設定方法の一つに，平均超過関数を用いる方法がある．以下に，平均超過関数について説明する．
確率変数 Y が尺度母数σと形状母数ξを有する一般化パレート分布に従うとき，確率変数 Y の期待値は以下のように得られる．

$$E(Y) = \frac{\sigma}{1-\xi}, \qquad \xi < 1 \tag{3.3.1.1}$$

ここで，確率変数 X が閾値 u_0 に対して一般化パレート分布に従うと仮定すると，対応する尺度母数 σ_{u_0} を用いると式(3.3.1.1)は以下のように表すことができる．

$$E\left(X - u_0 \middle| X > u_0\right) = \frac{\sigma_{u_0}}{1-\xi}, \quad \xi < 1 \tag{3.3.1.2}$$

さらに，確率変数 X が閾値 u_0 に対して一般化パレート分布に従えば，$u > u_0$ を満たす全ての閾値に対して，対応する尺度母数σ_u を有する一般化パレート分布にも従う．これは下記のように表すことができる．

$$\begin{aligned} E\left(X - u \middle| X > u\right) &= \frac{\sigma_u}{1-\xi} \\ &= \frac{\sigma_{u_0} + \xi u}{1-\xi} \end{aligned} \tag{3.3.1.3}$$

式(3.3.1.3)は，$u > u_0$を満たすuに対して E($X - u \mid X > u$)はuの線形関数であることを示している．また，E($X - u \mid X > u$)を平均超過関数と呼ぶ．観測されたデータから得られる標本平均超過関数は，以下の式で求めることができる．

$$E\left(X-u\middle|X>u\right)=\frac{1}{n_u}\sum_{i=1}^{n_u}\left(x_{(i)}-u\right),\qquad u<x_{\max} \tag{3.3.1.4}$$

ただしn_uは、$x>u$となるデータの個数である。
閾値の設定に関しては，式(3.3.1.4)で示した標本平均超過関数を閾値 u に対してプロットし，線形性が保たれている部分に着目すれば適切な閾値を設定することができる．

また，超過数を自動的に選択する以下の方法は、経験的にかなり合理的な超過数を選択できることが知られている(Reiss and Thomas, 1991; p.121)．今ξ_iを、I 個のデータに基づいて推定さえた形状母数、またその中央値を med (ξ_1,…, ξ_k)で表したとする．このとき，次のような k（すなわち k^*）を選択する．

$$k^*=\frac{1}{k}\sum_{i\le k} i^{\beta}\left|\xi_i-med\left(\xi_1,\cdots,\xi_k\right)\right|$$

ただし、$0\le\beta\le 1/2$ である。サンプル数が比較的少ないときは，一連の推定値を平滑化すると良い結果が得られる．この方法を若干修正したものに中央値の代わりにξ_k を用い，絶対偏差の代わりに二乗偏差を用いる方法もある．

3.3.2 再現レベル

実際のデータ解析では、解析結果を求められた確率分布のパラメータ値で示すよりも、再現値で示す場合が多い。これにより、いくつかの解析結果の比較も容易になる。

確率変数 X が閾値 u に対して，尺度母数σ，形状母数ξを有する一般化パレート分布に従うとき，$x > u$ に対して以下の関係が成り立つ．

$$\Pr\left\{X>x\middle|X>u\right\}=\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}} \tag{3.3.2.1}$$

式(3.3.2.1)は以下のように表すこともできる．

$$\Pr\left\{X>x\right\}=\zeta_u\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}} \tag{3.3.2.2}$$

ここで，$\zeta_u = \Pr\{X > u\}$である．平均的に観測数 m ごとに 1 度だけ閾値を超過するレベルを x_m とすると，式(3.3.2.2)から，

$$\zeta_u\left[1+\xi\left(\frac{x_m-u}{\sigma}\right)\right]^{-\frac{1}{\xi}}=\frac{1}{m} \tag{3.3.2.3}$$

式(3.3.2.3)を x_m について求めると，

$$x_m=u+\frac{\sigma}{\xi}\left\{\left(m\zeta_u\right)^{\xi}-1\right\} \tag{3.3.2.4}$$

ここで，$x_m > u$ を満たすように m は十分大きな値とする．式(3.3.2.1)から式(3.3.2.4)のξは$\xi \neq 0$ である．$\xi = 0$ の場合では，式(3.3.2.4)は以下のように表すことができる．

$$x_m=u+\sigma\log\left(m\zeta_u\right) \tag{3.3.2.5}$$

一般的には，確率量として年確率量を指標に用いる場合が少なくない．仮に，1 年における観測数が n_y であるとすると，式(3.3.2.4)の観測数 m は，$N \times n_y$ として求めることができる．よって，N 年確率量は，以下のように求めることができる．

$$z_N = u + \frac{\sigma}{\xi}\left\{\left(Nn_y\zeta_u\right)^{\xi} - 1\right\}, \quad \xi \neq 0 \tag{3.3.2.6}$$

$$z_N = u + \sigma\log\left(Nn_y\zeta_u\right), \quad \xi = 0 \tag{3.3.2.7}$$

ζ_u の推定値は以下のように求めることができる．

$$\hat{\zeta}_u = \frac{k}{N} \tag{3.3.2.8}$$

ただし N は、全データ数、k は閾値 u より大きいデータの数である。

3.3.3 モデルの妥当性の確認

モデルの妥当性を確認するためには，Probability plot や Quantile plot などがあるのは、極値分布解析の場合と同じである．Probability plot や Quantile plot が線形であれば推定したモデルは観測データによく当てはまっていると言える．閾値 u に対する超過変数を $y_{(1)} \le \cdots \le y_{(k)}$とし，推定モデルを $\hat{H}$ とすると，Probability plot は以下の組み合わせで作成することができる．

$$\left\{\left(\frac{i}{k+1}, \hat{H}\left(y_{(i)}\right)\right); i = 1,\ldots,k\right\} \tag{3.3.3.1}$$

ここで，推定モデル $\hat{H}$ は以下のように求めることができる．

$$\hat{H}(y) = 1 - \left(1 + \frac{\hat{\xi}y}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\xi}}} \tag{3.3.3.2}$$

また，Quantile plot は以下の組み合わせで作成することができる．

$$\left\{\left(\hat{H}^{-1}\left(\frac{i}{k+1}\right), y_{(i)}\right); i = 1,\ldots,k\right\} \tag{3.3.3.3}$$

ここで，$\hat{H}^{-1}$ は以下のように求めることができる．

$$\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(y^{-\hat{\xi}} - 1\right) \tag{3.3.3.4}$$

4. 母数の推定方法

4.1 最尤法

一般的に用いられている母数推定法の一つに最尤法がある．最尤法で用いる尤度関数は以下のように定義される．

$$L(\theta) = \prod_{i=1}^{n} f\left(x_i;\theta\right) \tag{4.1.1}$$

ここで，L は尤度，θ は母数，$f(x_i;\theta)$は確率変数 x と母数θを有する確率密度関数である．実際の計算では，式(4.1.1)で示した尤度の対数（対数尤度）を最大化したほうが都合がよい．対数尤度は以下のように表すことができる．

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta) \tag{4.1.2}$$

ここで，lは対数尤度である． 式(4.1.2)を最大化する一般的な方法として，Newton-Raphson 法が挙げられる．ただしこの手法は統計データが多峰性であった場合には，初期値によっては局所解に収束する場合があるので，初期値を変更して再度計算するなど注意する必要がある．

4.2 積率法

積率法は、理論的に計算された確率密度関数のいくつかのモーメント値（1 次モーメントは平均、2 次モーメントは分散等）と、データから計算されるモーメントを比較することにより、パラメータ推定を行おうという考え方であり、最尤法と並んで、統計学ではもっとも一般的な推定法である。例えば、Gumble 分布のパラメータ a,b と、標本平均 x と分散 s2 の間には、次の関係がある。

$$a = 6^{1/2}\, s_N / \pi \tag{4.2.1}$$

$$b = \bar{x} - a\lambda \tag{4.2.2}$$

ただし、l はオイラー数で 0.577216・・・であり、xは標本平均、s_Nは標本標準偏差を表す。
モーメント法は、統計解析では、広く用いられており、極値統計学も例外ではない。しかし、この方法による推定値は、不安定な場合が多く、次に述べるようないろいろな工夫が提案されている。

4.3 PWM 法

Probability Weighted Moment (PWM) 法は 1979 年 Greenwood らにより提案された確率分布の母数推定法である．この手法は，確率分布の母数推定の規準として通常の積率に替えて，確率で加重した積率を用いるという手法である．この手法により，標本の（高次）モーメントを用いることによる積率法の難点を標本そのものではなくその非超過確率のモーメントを用いることで回避している．PWM は以下のように定義される．

$$M_{l,j,k} = E\left[X^l F^j (1-F)^k\right] = \int_0^1 x^l F^j (1-F)^k \, dF \tag{4.3.1}$$

ここで，$M_{l,j,k}$は PWM，l, j, kは非負の整数，Fは確率分布関数である．Fに対象とする確率分布関数を代入すれば，l, j, kの定まった組み合わせに対し式(4.3.1)の積分を実行することができる．積分結果はFの母数の関数になっている．したがって，未知母数に等しい数のl, j, kの組み合わせに対する PWM を求めれば，未知母数に関する連立方程式が得られ，その解として母数を求めることができる．上記方法を実行するためには，まずl, j, kの組み合わせを決めなくてはならない．PWM 法では，$l = 1, j = 0$もしくは$l = 1, k = 0$のどちらか都合のよい方を用いる．水文頻度解析においては，$l = 1, k = 0$がよく用いられている?)．$l = 1, k = 0$とした場合の PWM は次式で定義される．

$$M_{1,j,0} = M_j = E\left[XF^j\right] = \int_0^1 xF^j dF \tag{4.3.2}$$

M_jの標本推定値算定には 2 つの方法がある．もっとも単純な方法はx_iの生起確率に適当なプロッティング・ポジションF_iを仮定する以下の方法である．

$$\hat{M}_j = \frac{1}{N}\sum_{i=1}^{N} x_i F_i^j \tag{4.3.3}$$

この場合には，プロッティング・ポジション公式に何を選択するかが重要な問題となる．Hosiking ら(1985)は GEV 分布のクオンタイル推定に下記の公式を用いており，水文頻度解析では広く用いられている．

$$F_i = \frac{i - 0.35}{N} \tag{4.3.4}$$

理論的に M_j の不偏推定値 $\hat{M}_j$ は以下のように算定できる．

$j = 0$ の場合，

$$\hat{M}_j = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4.3.5}$$

$j \geq 0$ の場合，

$$\hat{M}_j = \frac{1}{N}\sum_{i=1}^{N} x_i \frac{(i-1)(i-2)\cdots(i-j)}{(N-1)(N-2)\cdots(N-j)} \tag{4.3.6}$$

4.4 L 積率法

L 積率法は，1990 年に Hosking によって提案された確率分布の母数推定法である．L 積率は，通常用いられている積率と異なり，順序統計量の線形和で表される特徴をもつ． L 積率は線形和に基づく統計量であるため，小標本から計算される変動係数やひずみ係数は偏りが大で変動も大きいという問題を緩和させるためにも有効である．第 1 次の L 積率は次式で定義される．

$$\lambda_1 = E(X) \tag{4.4.1}$$

ここに，$E(\cdot)$は期待値演算子である．いま，$X_{(i|N)}$を大きさ N の標本の第 i 番目の順序統計量とするとき，第 2 次の L 積率は任意に 2 個取り出したときの最大値と最小値の差の期待値に比例して，次式で定義される．

$$\lambda_2 = \frac{1}{2}E\left[X_{(1|2)} - X_{(2|2)}\right] \tag{4.4.2}$$

同様に，第 3 次と第 4 次の L 積率も以下のように定義される．

$$\lambda_3 = \frac{1}{3}E\left[X_{(1|3)} - 2X_{(2|3)} + X_{(3|3)}\right] \tag{4.4.3}$$

$$\lambda_4 = \frac{1}{4}E\left[X_{(1|4)} - 3X_{(2|4)} + 3X_{(3|4)} - X_{(4|4)}\right] \tag{4.4.4}$$

L 積率と PWM は次式で関係づけられる．

$$\lambda_1 = \hat{M}_0 \qquad \text{（第 1 次 L 積率）} \tag{4.4.5}$$

$$\lambda_2 = 2\hat{M}_1 - \hat{M}_0 \qquad \text{（第 2 次 L 積率）} \tag{4.4.6}$$

$$\lambda_3 = 6\hat{M}_2 - 6\hat{M}_1 + \hat{M}_0 \qquad \text{（第 3 次 L 積率）} \tag{4.4.7}$$

PWM の M_j ($j = 0, 1, 2, \cdots$)の標本推定値が得られていれば，L 積率の推定値を容易に算定することができる．また，分布母数と PWM ないし L 積率との関係式が得られていれば，母数推定のための L 積率解が求められる．

5. 確率分布モデルと推定母数

5.1 極値分布

5.1.1 ガンベル分布

ガンベル分布の確率分布関数 F_X と確率密度関数 f_X は，それぞれ以下のように表される．

$$F(x)=\exp\left[-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right] \tag{5.1.1.1}$$

$$f_X(x)=\frac{1}{\sigma}\exp\left[-\frac{x-\mu}{\sigma}-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right] \tag{5.1.1.2}$$

ガンベル分布のクオンタイル（式(5.1.1.1)の逆関数）は，次式により容易に算出される．

$$x_p=\mu-\sigma\ln\left[-\ln(F)\right] \tag{5.1.1.3}$$

ガンベル分布の PWM と L 積率解は，次式で与えられる．

$$\mu=\hat{M}_0-0.5772\frac{2\hat{M}_1-\hat{M}_0}{\ln 2}=\lambda_1-0.5772\frac{\lambda_2}{\ln 2} \tag{5.1.1.4}$$

$$\sigma=\frac{2\hat{M}_1-\hat{M}_0}{\ln 2}=\frac{\lambda_2}{\ln 2} \tag{5.1.1.5}$$

5.1.2 一般化極値分布（GEV 分布）

先に紹介したように Jenkinson (1955)は 3 種の極値分布を 1 つの式形に統一して，一般化極値分布の導入を図った．イギリスの Natural Environmental Research Council は年最大日流量の確率モデルにつき検討を行い，GEV 分布を基準法として推奨している（Natural Environmental Research Council: Flood Studies Report 1975）．一般化極値分布の確率分布関数 F_X と確率密度関数 f_X は，それぞれ以下のように表される．

$\xi\neq 0$ の場合，

$$F(x)=\exp\left[-\left\{1+\xi\left(\frac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\right] \tag{5.1.2.1}$$

$$f(x)=\frac{1}{\sigma}\cdot\left\{\frac{\sigma+\xi\cdot(x-\mu)}{\sigma}\right\}^{-\frac{1}{\xi}-1}\cdot\exp\left\{-\left(\frac{\sigma+\xi\cdot(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right\} \tag{5.1.2.2}$$

となり，$\xi>0$ の場合には Fréchet 分布に，$\xi<0$ の場合には Weibul 分布にそれぞれ一致する．$\xi=0$ の場合，

$$F(x)=\exp\left[-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right] \tag{5.1.2.3}$$

$$f_X(x)=\frac{1}{\sigma}\exp\left[-\frac{x-\mu}{\sigma}-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right] \tag{5.1.2.4}$$

となり，Gumbel 分布に一致する．GEV 分布のクオンタイル（式(3.5.4)の逆関数）は，次式により容易に算出される．

$$x_p = \mu - \frac{\sigma}{\xi}\left[1 - \left\{-\ln(F)\right\}^{-\xi}\right] \quad (\xi \neq 0) \tag{5.1.2.5}$$

$$x_p = \mu - \sigma \ln\left[-\ln(F)\right] \quad (\xi = 0) \tag{5.1.2.6}$$

GEV 分布の PWM 解は次式で与えられる．

$$\hat{M}_j = (j+1)^{-1}\left[\mu - \frac{\sigma}{\xi}\left\{1 - \frac{\Gamma(1-\xi)}{(j+1)^{-\xi}}\right\}\right] \quad (\xi < 1,\ j = 0,1,2,\cdots) \tag{5.1.2.7}$$

したがって，L 積率は次式で与えられる．

$$\lambda_1 = \hat{M}_0 = \mu - \frac{\sigma}{\xi}\left\{1 - \Gamma(1-\xi)\right\} \tag{5.1.2.8}$$

$$\lambda_2 = 2\hat{M}_1 - \hat{M}_0 = -\frac{\sigma}{\xi}\left(1 - 2^{\xi}\right)\Gamma(1-\xi) \tag{5.1.2.9}$$

$$\frac{2\lambda_2}{\lambda_3 + 3\lambda_2} = \frac{2\hat{M}_1 - \hat{M}_0}{3\hat{M}_2 - \hat{M}_0} = \frac{1 - 2^{\xi}}{1 - 3^{\xi}} \tag{5.1.2.10}$$

GEV 分布の L 積率解を得るためには式(5.1.2.10)をξ について解く必要がある．この近似解は，Hosking ら(1985)により以下のように提案されている．

$$\xi \approx 7.8590d + 2.9554d^2 \tag{5.1.2.11}$$

ここに，式(5.1.2.11)の d は以下のように与えられる．

$$d = \frac{2\hat{M}_1 - \hat{M}_0}{3\hat{M}_2 - \hat{M}_0} - \frac{\ln 2}{\ln 3} = \frac{2\lambda_2}{\lambda_3 + 3\lambda_2} - \frac{\ln 2}{\ln 3} \tag{5.1.2.12}$$

近似式(5.1.2.12)は，式(5.1.2.10)に対して$-1/2 < \xi < 1/2$ の範囲で最大誤差は 0.0009 以下であり，実用上十分な精度を有している．形状母数ξ を求めた後，式(5.1.2.8)と式(5.1.2.9)を用いることで位置母数μ と尺度母数σ の PWM と L 積率解を以下のように求めることができる．

$$\mu = \hat{M}_0 - \frac{\left(\hat{M}_0 - 2\hat{M}_1\right)\left(1 - \Gamma(1-\xi)\right)}{\left(2^{\xi} - 1\right)\Gamma(1-\xi)} = \lambda_1 + \frac{\lambda_2\left(1 + \xi\Gamma(-\xi)\right)}{\left(2^{\xi} - 1\right)\Gamma(1-\xi)} \tag{5.1.2.13}$$

$$\sigma = -\frac{\hat{M}_0 - 2\hat{M}_1}{\left(2^{\xi} - 1\right)\Gamma(-\xi)} = -\frac{\lambda_2}{\left(2^{\xi} - 1\right)\Gamma(-\xi)} \tag{5.1.2.14}$$

5.2 閾値モデル

5.2.1 指数分布

指数分布の確率分布関数 F_X と確率密度関数 f_X は，それぞれ以下のように表される．

$$F(x) = 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) \tag{5.2.1.1}$$

$$f(x) = \frac{1}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\} \tag{5.2.1.2}$$

指数分布のクオンタイル（式(5.2.1.1)の逆関数）は，次式により容易に算出される．

$$x = \mu - \sigma \ln(1 - F) \tag{5.2.1.3}$$

指数分布の PWM と L 積率解は，次式で与えられる．

$$\mu = -\hat{M}_0 - 4\hat{M}_1 = \lambda_1 - 2\lambda_2 \tag{5.2.1.4}$$

$$\sigma = 2(2\hat{M}_1 - \hat{M}_0) = 2\lambda_2 \tag{5.2.1.5}$$

5.2.2 一般化パレート分布

一般化パレート分布の確率分布関数 F_X と確率密度関数 f_X は，それぞれ以下のように表される．

$$F(x) = 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{5.2.2.1}$$

$$f(x) = \frac{1}{\tilde{\sigma}}\left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-\frac{1+\xi}{\xi}} \tag{5.2.2.2}$$

ここで，$\xi \neq 0$，$x > 0$，$1 + \frac{\xi x}{\tilde{\sigma}} > 0$ であり，$\tilde{\sigma}$ は下記のように求められる．

$$\tilde{\sigma} = \sigma + \xi(u - \mu) \tag{5.2.2.3}$$

$\xi = 0$ の場合では，

$$F(x) = 1 - \exp\left(-\frac{x}{\tilde{\sigma}}\right) \tag{5.2.2.4}$$

$$f(x) = \frac{1}{\tilde{\sigma}} \exp\left\{-\frac{x}{\tilde{\sigma}}\right\} \tag{5.2.2.5}$$

となり，指数分布となる．ここで，式(5.2.2.1)から式(5.2.2.3)中のμ，σ，ξは GEV 分布の母数と一致する．一般化パレート分布のクオンタイルは次式で表される．

$$x = \frac{\tilde{\sigma}}{\xi}\left\{(1 - F)^{-\xi} - 1\right\} \tag{5.2.2.6}$$

一般化パレート分布の PWM と L 積率解は，次式で与えられる．

$$\xi = \frac{3M_0 - 4M_1}{M_0 - 2M_1} = 2 - \frac{\lambda_1}{\lambda_2} \tag{5.2.2.7}$$

$$\tilde{\sigma} = -\frac{2M_0(M_0 - M_1)}{M_0 - 2M_1} = \frac{\lambda_1(\lambda_1 - \lambda_2)}{\lambda_2} \tag{5.2.2.8}$$

5.3 その他の確率分布モデル

水文頻度解析では，正規分布，対数正規分布，平方根指数型最大値分布，ピアソンⅢ型分布，対数ピアソンⅢ型分布，一般化ガンマ分布などが用いられている．これらの分布の概要は，文献 14 を参考にされたい．

6. 確率量の誤差推定

6.1 デルタ法

6.2 jackknife 法

確率量の誤差推定を行う手法の一つに jacknife 法がある．まず，N 個のデータ $x_1, x_2, \cdots, x_N$ を用いてその母集団の特性を表す量を推定する統計量ををφと記す．推定誤差は以下のように求められる．

$$\hat{s}_J^2 = \frac{N-1}{N}\sum_{i=1}^{N}\left(\varphi_i - \varphi_{ave}\right)^2 \tag{6.2.1}$$

ここで，$\hat{s}_J^2$ は jacknife 法による推定誤差，N はデータの個数，φ_i は i 番目のデータを除いた $N-1$ 個のデータを用いた統計量であり，以下のように算出する．

$$\varphi_i = \varphi\left(x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_N\right) \tag{6.2.2}$$

φ_{ave} は φ_i の平均であり，以下のように求める．

$$\varphi_{ave} = \frac{1}{N}\sum_{i=1}^{N}\varphi_i \tag{6.2.3}$$

jackknife 法を用いた検討事例は，寶・高棹(1988)などがある．また，1974 年に Miller によって jackknife 法に関する詳細なレビューがなされている．

6.3 boostrap 法

boostrap 法は 1979 年に Efron により定式化されたリサンプリング法である．まず，N 個のデータ $x_1, x_2, \cdots, x_N$ から繰返しを許して N 個取り出し，それを $x_1^*, x_2^*, \cdots, x_N^*$ と記す．この 1 組のデータを boostrap 標本という．boostrap 標本を用いて統計量を求め，それを次のように記す．

$$\hat{\varphi}^* = \varphi\left(x_1^*, x_2^*, \cdots, x_N^*\right) \tag{6.3.1}$$

統計量φの boostrap 推定値は，以下のよう求められる．

$$\hat{\varphi}_{ave}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\varphi}^{*b} \tag{6.3.2}$$

ここで，$\hat{\varphi}^{*b}$ は，第 b 番目の boostrap 標本に対して得られた統計量である．統計量φの分散の boostrap 推定値は，以下のよう求められる．

$$\hat{s}_B^2 = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\varphi}^{*b} - \hat{\varphi}_{ave}^*\right)^2 \tag{6.3.3}$$

7. 例題

7.1 東京における地震動の極値分布

宇佐美カタログにより，福島・田中の距離減衰式を用いて東京都庁所在地（北緯 35 度 41 分、東経 139 度 45 分）における 1600 年から 1995 年までの地震の最大加速度を推定した． この結果を**図 8.1.1** に示した． なお、震源の深度が欠落しているデータに関しては一律に 30km を仮定した．用いた福島・田中の距離減衰式は、次の通りである．

$$\log A = 0.51M - \log\{R + 0.006\exp(1.174M)\} - 0.0033R + 0.37 + 0.22S$$

ここに，A = 地表面最大加速度 (以下の式では x により示す。)， M = マグニチュード， R = 震央距離， S = 地盤種別を表すパラメータ: S = 1 土、S = 0 岩盤; この例題では S = 1 が用いられた．

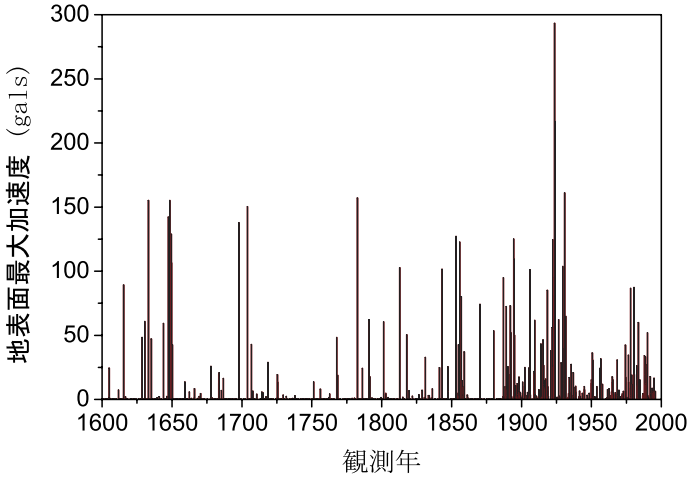

図 7.1.1　1600 年から 1995 年までの東京都庁地点での地表面最大加速度
（福島・田中式より推定）

観測データの大きい方から k 番目までのデータを用いて POT 解析を行い、L 積率法により母数の推定を行った．形状母数を k に対してプロットしたのが，図 7.1.2 である．データ数が少ない範囲では，推定値は不安定であり，データ数が 50 個から 70 個の範囲では，形状母数の推定値は安定した値が得られることが分かる．図 7.1.3 は平均超過関数である．図 7.1.3 から閾値が小さい場合（データ数が多い場合）では，平均超過関数は増加傾向にあり，閾値が 30gal 付近（データ数約 70 個）を境にして平均超過関数は減少傾向に転じる．さらに，閾値が大きい場合（150 gal を超えた場合）では，データ数が少なくなることもあり，平均超過関数は閾値が小さい場合と比較して大きく異なる．図 7.1.4 と図 7.1.5 は再現期間を 100 年とした場合のデータ数と閾値に対応する確率地表面最大加速度である．全体的に，設定したデータ数や閾値により確率地表面最大加速度が大きく異なることが分かる．図 7.1.2 の結果と同様にデータ数が 50 個から 70 個の範囲では，安定した確率地表面最大加速度が得られる．また，図 7.1.5 から閾値が 30 gal から 100 gal 付近までは安定した確率地表面最大加速度が得られる．以上の結果から，本例題では閾値を 35 gal，データ数を 60 個に設定した．

表 7.1.1 に閾値モデルによる解析結果を示す．モデルの妥当性を検証するために，Probability Plot と Quatile Plot を行った（**図 7.1.6** と**図 7.1.7**）．Probability Plot と Quatile Plot の両方とも直線性が保たれており，当てはまりの適切さを示している．

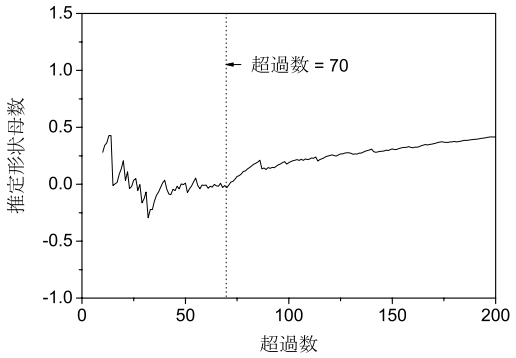

図 7.1.2　データ数と形状母数の関係

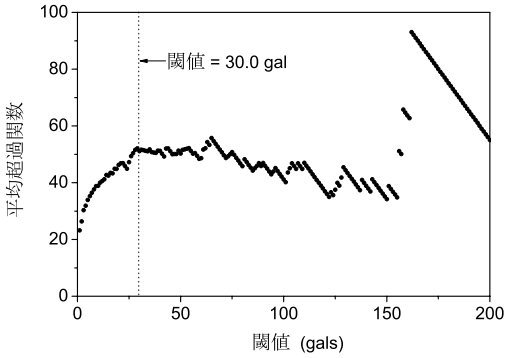

図 7.1.3　平均超過関数

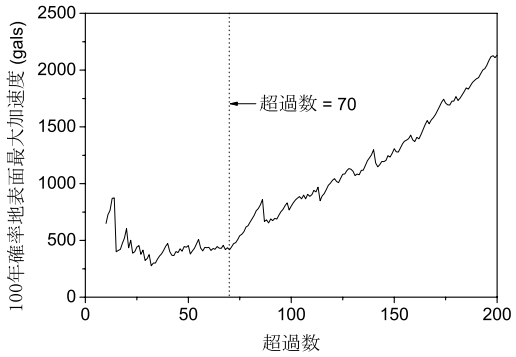

図 7.1.4　データ数と 100 年確率最大地表面加速度の関係

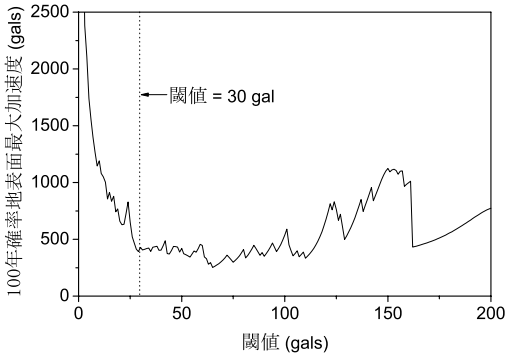

図 7.1.5　閾値と 100 年確率最大地表面加速度の関係

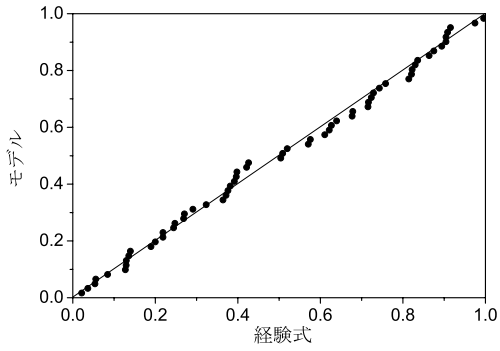

図 7.1.6　Probability Plot

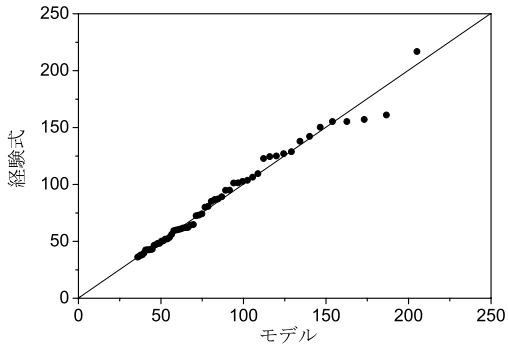

図 7.1.7　Quantile Plot

表 7.1.1　閾値分布モデルによる統計解析結果

位置母数	尺度母数	形状母数	100 年確率量 gal
35.0	54.60	-0.055	393.5

7.2 東京における日降水量の極値分布

（財）気象事業支援センターから提供されている地域気象観測所（AMeDAS）の統計量のうち，観測点東京（北緯 35 度 41 分，東経 139 度 45 分）における日降水量の統計解析を実施した．統計解析は極値分布モデルと閾値モデルを用いた．

7.2.1 極値分布モデルによる推定

極値分布モデルでは，1980 年から 2004 年までの年ごとの最大日降水量を基に統計解析を実施した．データ数は合計 29 個である．Gumbel 分布と一般化極値分布を用いて，それぞれの確率分布に対して，L 積率法により推定した母数，確率最大日降水量，確率降水量の変動係数の jackknife 推定値を示した． 表 7.2.1.1 から推定誤差がほぼ一致していること，一般化極値分布を用いても形状母数が 0 に近いことから，Gumbel 分布が適していることが分かる．

表 7.2.1.1　極値分布モデルによる統計解析結果

確率分布	位置母数	尺度母数	形状母数	100 年確率量 mm/hour	推定誤差
Gumbel 分布	105.03	48.75	0.0	329.3	0.095
一般化極値分布	105.03	48.75	0.0	329.3	0.080

モデルの妥当性を検証するために，Probability Plot と Quatile Plot を行った（**図 7.2.1.1** と**図 7.2.1.2**）．Probability Plot と Quatile Plot の両方とも直線性が保たれており，当てはまりの適切さを示している．

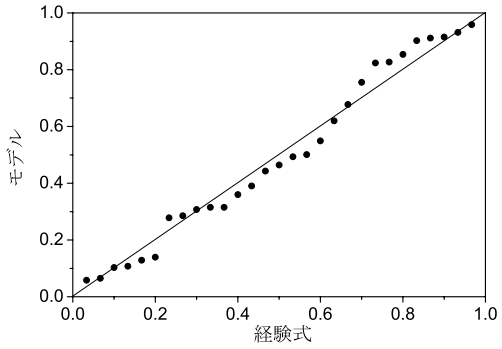

図 7.2.1.2　Probability Plot

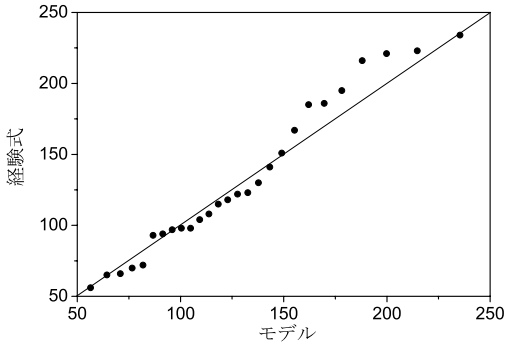

図 7.2.2.2　Quantile Plot

7.2.2 閾値モデルによる推定

閾値モデルでは，1980 年から 2004 年までの日降水量として合計 9131 個のデータを用いた（図 7.2.2.1）．これらのデータには降水量がゼロのデータも含まれている．

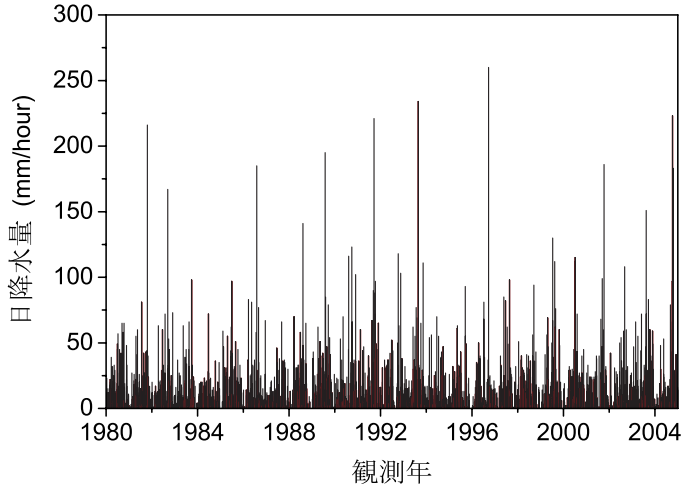

図 7.2.2.1　1980 年から 2004 年における観測点東京での日降水量

観測データの大きい方から k 番目までのデータを用いて POT 解析を行い、L 積率法により母数の推定を行った．形状母数を k に対してプロットしたのが，**図 7.2.2.2** である．データ数が少ない範囲では，推定値は不安定であり，データ数が 120 個以上では，形状母数の推定値は安定した値が得られることが分かる．**図 7.2.2.3** は平均超過関数である．**図 7.2.2.3** から，閾値が 50 mm/hour 付近（データ数約 120 個）までは平均超過関数はほぼ線形であることが分かる．**図 7.2.2.4** と**図 7.2.2.5** は再現期間を 100 年とした場合のデータ数と閾値に対応する確率最大日降水量である．全体的に，設定したデータ数や閾値により確率最大日降水量が大きく異なることが分かる．**図 7.2.2.2** の結果と同様にデータ数が 120 個以下では確率最大日降水量の変動が大きく，データ数が 120 個から 200 個付近で安定した確率最大日降水量が得られることが分かる．また，**図 7.2.2.5** から閾値が 50 mm/hour を超えると確率最大日降水量の変動が大きくなることが分かる．以上の結果から，本例題では閾値を 50 mm/hour，データ数を 120 個に設定した結果，**表 7.2.1.1** に示す推定母数と 100 年確率量が得られた．

閾値モデルの妥当性を検証するために，Probability Plot と Quatile Plot を行った（**図 7.2.2.6** と**図 7.2.2.7**）．Probability Plot と Quatile Plot の両方とも直線性が保たれており，当てはまりの適切さを示している．**表 7.2.2.2** では，極値分布モデルと閾値モデルから得られた形状母数と 100 年確率量を比較した．前述したように，極値分布モデルと閾値モデルにおける形状母数は理論的に一致するが，本例題では若干異なる結果となった．これは，極値分布モデルで用いた統計データが少なかったことが要因の一つに挙げられる．

表 7.2.1.1　閾値分布モデルによる統計解析結果

位置母数	尺度母数	形状母数	100 年確率量 mm/hour
50.0	27.20	0.2179	404.4

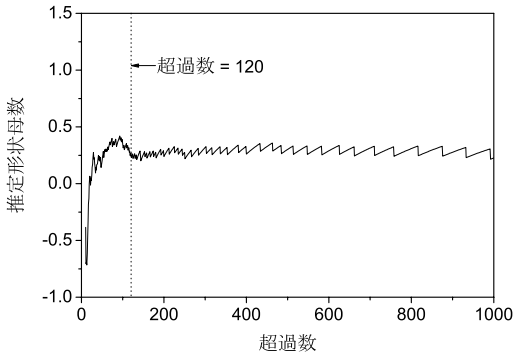

図 7.2.2.2　データ数と形状母数の関係

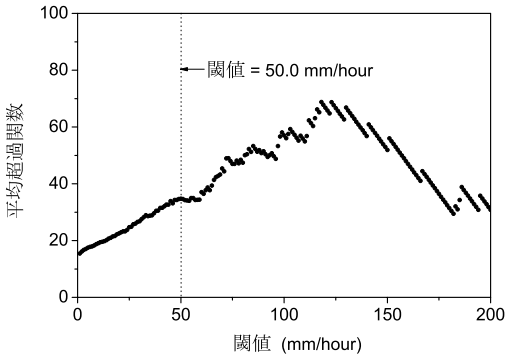

図 7.2.2.3　平均超過関数

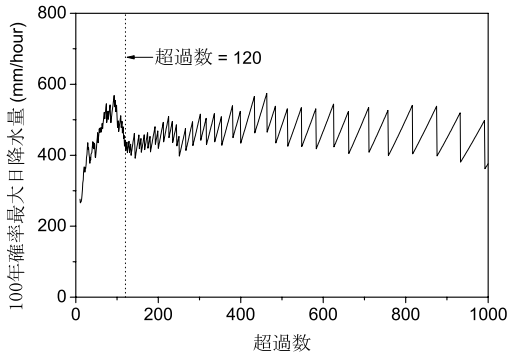

図 7.2.2.4　データ数と 100 年確率最大日降水量の関係

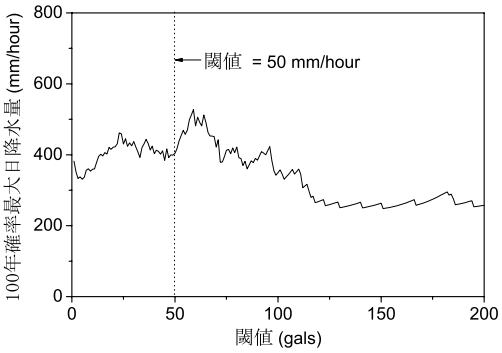

図 7.2.2.5　閾値と 100 年確率最大日降水量の関係

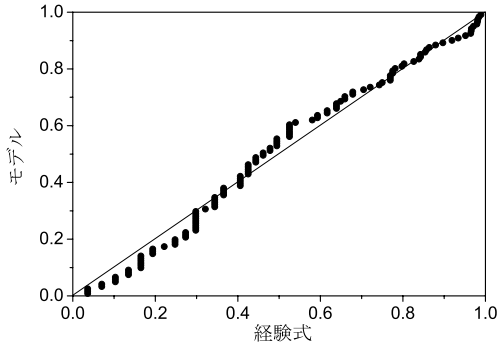

図 7.2.2.6　Probability Plot

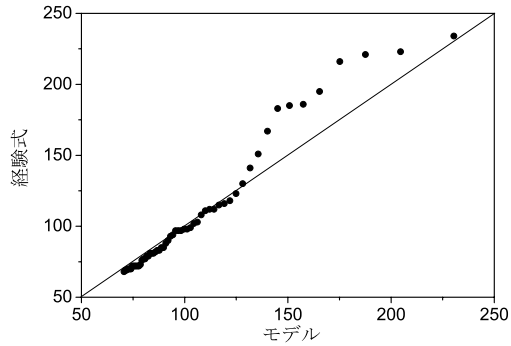

図 7.2.2.7　Quantile Plot

表 7.2.2.2　極値分布モデルと閾値モデルの比較

モデル	データ数	形状母数	100 年確率量 mm/hour
極値分布モデル	29	0.0	329.3
閾値モデル	120	0.2179	404.4

8. 計算プログラム

極値分布モデルや閾値モデルを用いた統計解析を行うためには，計算ツールが必須となる．下記は，代表的な統計解析プログラムである．使用にあたっては，理論的背景を十分に理解することが重要である．

表 8.1　代表的な統計解析プログラム

作成者	入手先
Reiss, R.D. and Thomas, M.	http://www.risktec.de/index.htm
Heffernan, J.	http://www.maths.lancs.ac.uk/~currie/index.html
Yee, T.	http://www.stat.auckland.ac.nz/~yee/
Hosking, J. R. M.	http://www.research.ibm.com/people/h/hosking/lmoments.html

9. むすび

極値の統計解析の研究は、広く普及している極値分布に基づいて確率紙等を用いて データのこれら分布への当てはめを行う 1960 年代に Gumbel により確立され、その後広く普及した方法から、一般パレート分布の発見による POT 解析への 大きく姿を変えてきている。特に金融工学に関連した現実問題の解決が求められ、これを受けて実用的なプログラムもすでに開発、配布されている。このような新しい成果が、土木工学の分野でも応用される必要がある。しかしこのために解決しなければならない問題も多い。特に確率点(quantile)の 信頼性評価の問題は重要であると思われる。

参考文献

[1] 高橋倫也(1999)「極値統計学へのお誘い」,日本統計学会会報, 1999 年 12 月号

[2] Fisher, R. A. and Tippettm, L. H. C. (1928): On the estimation of the frequency distributions of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, Vol. 24, pp. 180-190.

[3] von Mises(1936)

[4] Gnedenko, B. V. (1943): Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Statistics*, Vol. 44, pp. 423-453.

[5] Gumbel, E. J. (1958): Statistics of Extremes, Columbia University Press, New York.

[6] Weibull(1951)

[7] Balkeman & de Haan(1974)

[8] Pickands, J. (1975): Statistical inference using extreme order statistics, *Annals of Statistics*, Vol. 3, pp. 119-131.

[9] Teugels，1999

[10] Smith(1989)

[11] de Haan, 1999

[12] 日本建築学会(2004)：建築物荷重指針・同解説、pp.109-115.

[13] 土木学会・海岸工学委員会・研究現況レビュー小委員会(2001)：新しい波浪算定法とこれからの海域施設の設計法、pp.65-84.

[14] 水文・水資源学会(1997)：水文・水資源ハンドブック、pp.228-255.

[15] Coles, S. (2001) An introduction to statistical modeling of extreme values, Springer, pp.1-91.

[16] Jenkinson (1955)

[17] Reiss, R.D. and Thomas, M. (1997): Statistical analysis of extreme values, Birkhauser, pp316

[18] Greenwood, J. A., Landwehr, J. M. and Matalas, N. C. (1979): Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form, Water Resources Research, Vol. 15, No.5, pp.1049-1054.
[19] Hosking, J. R. M. (1990): L-Moments; analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistics Society*, B, 52, 2, pp.105-124.
[20] Jenkinson, A. F. (1955): The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteor. Soc.*, Vol. 81, pp. 158-171
[21] Natural Environmental Research Council (1975): Flood Studies Report, Vol. 1, Hydrological Studies, pp.550.
[22] Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985): Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics*, Vol. 27, No. 3, pp.251-261.
[23] 星清，豊沢康男 (1979): 一般化ガンマ分布の水文統計への適用，土木学会北海道支部論文報告集, No. 35, pp. 180-185.
[24] 宝馨, 高棹琢馬(1988): 水文頻度解析における確率分布モデルの評価基準, 土木学会論文集, No.393/II-9, pp.151-160.
[25] Miller, R. G. (1974): The jackknife—a review, Biometrica, Vol. 61, No. 1, pp. 1-15.
[26] Efron, B. (1979): Boostrap method: another look at the jackknife. Ann. Statist., Vol. 7, pp.1-26.