I-403　極値II型分布の特性とその年最大風速資料に対する適用性について

日本テトラポッド株式会社　小野澤昌巳
横浜国立大学工学部　正員　合田良実

1．まえがき

毎年最大風速のデータは、設計風速を選定する際の基本資料として極値統計解析が行われる。その際の分布関数として極値I型分布が広く用いられるけれども、観測地点によってはこの分布にうまく適合しないものがあり、このためSimiu & Filliben[1)]や藤野ほか[2)]が極値II型分布のあてはめを併用している。しかし、極値II型分布の特性についての検討は未だ不十分と思われるので、Monte Carlo 法を活用してプロッティング・ポジション公式や確率風速の信頼区間推定式を導き、気象庁の全国128地点のデータに適用した結果をここに報告する。

2．極値II型分布の関数形

極値分布に用いられる関数は、一般に尺度母数Aと位置母数Bを含むのが普通である。分布関数によってはさらに形状母数kも取り込んでいる。極値II型分布に対する3母数表現としては、

$$F(x)=\exp\{-[1+(x-B)/kA]^{-k}\} \qquad (1)$$

の関数形をここに提案する。ここに、xは確率変量（風速値）、F（x）はxの分布関数である。この形であれば、$k\to\infty$のときに極値I型分布に漸近することが明瞭であり、Simiu & Filliben[1)]が提示したような間接証明を用いる必要がない。

3母数型分布関数に対して極値データをあてはめる場合、3個の母数を同時に推定する方法（最尤法その他）と、1個を固定して残りの2母数を推定する方法とがある。標本の大きさが数十程度のときは、一般に母数推定の信頼度が低いので、後者の方法を用いるべきであると考えられる。ここでは、1／kの値が等間隔になるように、k＝2.5, 10/3, 5.0, & 10.0 の4個の値に固定して検討を進めた。

3．プロッティング・ポジション公式の誘導

分布関数のあてはめ方法にはいろいろなものがあるが、最小2乗法がもっとも分かりやすくかつ信頼度が高いと考えられる。この方法の一つの難点は、プロッティング・ポジションの選択を誤ると確率風速の推定値に偏り（Bias）を生じることである。極値I型分布に対しては Gringorten 公式を用いるべきであって、Weibull 公式（Thomas公式と誤称されることが少なくない）を用いるとかなり大きな正の偏りをもたらす。これは著者の一人[3)]も例証しているところである。極値II型分布に最小2乗法を適用するための公式は未だ提案されていないので、数値実験結果に基づいて、試行錯誤により式（2）を導いた。

$$\hat{F}_i=[i-(0.44-0.41/k)]/[N+(0.12-0.11/k)] \qquad (2)$$

ここに、iは確率変量xを昇順に並べ替えたときの順位数（i＝1が最小値）、$\hat{F}_i$ は順序統計量x_iに割り付ける非超過確率である。

式（2）は、上述の4個の形状母数を与えた母関数から大きさがN＝10, 15, 20, 30, 40, 50, 60 & 100の標本各10,000組を疑似一様乱数を使用して作成し、各標本に対して分布関数のあてはめを行なって確率風速を推定し、母関数における真値との差が最小になるようにしたものである。確率風速推定の再現期間としては、標本の大きさの10倍のところについて作業を行なった。推定値の平均値と真値との差は-0.7%～0.2%である。なお、式(2) の形から明らかなように、$k\to\infty$では Gringorten 公式に一致する。

4．確率風速推定値の信頼区間

毎年最大風速のデータから 100年確率風速などを推定するためには、観測されたN年間のデータを大きさの順に並べ替えて順序統計量x_iとし、その各々に式（2）で割り付けられる非超過確率に基づいて規準化変量を$y_i=F^{-1}(\hat{F}_i)$として計算する。そして、$x_i=Ay_i+B$の関係を設定して最小2乗法を適用

し、母数ＡおよびＢを推定する。この作業では、形状母数ｋは前述の４個の値のいずれか一つを用いる。母数の値が推定されれば、再現期間Ｒ年に対応する確率風速が非超過確率Ｆ＝１－１／Ｒに対するｘの値として式（１）を解いて求められる。

確率風速推定の最大の問題は、観測データが一つの統計的標本に過ぎないため、母数の推定値が母集団の真値からかなりずれている可能性があることである。このため、確率風速の推定値もある幅の統計的誤差を伴っている。この誤差の大きさを評価するため、母関数毎にＮ＝10～100 の標本を各10,000組Monte Carlo法で抽出し、各標本毎に確率風速の推定を行なった。この推定値$\hat{x}_R$と真値x_Rの差を標本の不偏標準偏差σ_xで除して補助統計量ｚを作成し、このｚについて統計解析を行なってその標準偏差を計算した。この結果を実験式の形に取りまとめたのが次式である。

$$\sigma(\hat{x}_R) = [a + b\,y_R^2]^{1/2}\,\sigma_x / \sqrt{N} \qquad (3)$$

ここに、y_Rは再現期間Ｒに対応する基準化変量であり、次式で算定される。

$$y_R = k\,[-\ln(1-1/R)]^{-1/k} - 1 \qquad (4)$$

また、ａおよびｂは形状母数ｋの値ごとに次のように定式化した。

$$\begin{aligned} k &= 2.5: & a &= 0.2, & b &= 4.37\,\exp(3.5/N) \\ k &= 10/3: & a &= 0.4, & b &= 3.49 - \exp(2.0/N) \\ k &= 5.0: & a &= 0.6, & b &= 2.45 - \exp(1.5/N) \\ k &= 10.0: & a &= 0.9, & b &= 0.95 \end{aligned} \qquad (5)$$

なお、極値Ｉ型分布に対する確率風速の標準偏差を同様の手法で求めた結果は文献[3]に記載されている。

５．気象庁データに対するあてはめ結果

気象庁では先に1929～1966年間の日本各地 131地点の年最大風速に極値Ｉ型分布をあてはめた結果を発表している[4]。今回はこれに1987年までのデータを追加したが、筑波山、富士山、雲仙岳はその後の風速統計から除外されているので、残りの 128地点のデータを対象とした。データのあてはめには極値Ｉ型分布と上述の４種類の極値Ⅱ型分布を用い、x_iとy_iの間の相関係数の値が最大のものを最適合関数とみなした。極値Ⅱ型分布が最適合と判断された地点は次のとおりであり、その他は極値Ｉ型分布が最適合であった。

ｋ＝ 2.5： **大阪** (1)　　ｋ＝10/3： **岩見沢**、**甲府**、**福井**、**舞鶴** (4)
ｋ＝ 5.0： 松本、岐阜、名古屋、**彦根**、**京都**、鳥取、佐賀 (7)
ｋ＝10.0： 倶知安、室蘭、寿都、山形、水戸、銚子、東京、勝浦、八丈島、石廊崎、御前崎、軽井沢、尾鷲、和歌山、岡山、剣山、室戸岬、宿毛、松山、厳原、佐世保、熊本、人吉、阿久根、宮崎、種子島 (26)

なお、太字で示した７地点はデータ中の最大値が残りのものと非常にかけ離れているために、これらのデータが極値Ｉ型分布から抽出された可能性が５％未満と判断される。また、岩見沢、甲府、大阪の３地点は上記の可能性が１％未満である（この判断基準については文献[5]を参照されたい）。逆にいえば、太字でない地点は極値Ｉ型分布に所属する可能性が５％以上あることになる。

極値Ⅱ型分布の大きな問題点は、確率風速の推定誤差が非常に大きくなることである。たとえば、大阪のデータにｋ＝2.5 をあてはめて 100年確率風速を求めると36.0 m/s の値が得られるが、これに対する標準偏差を式（３）で計算すると18.3 m/s となり、実際問題としては採択することがむずかしい。むしろ、波浪の極値統計[3]で使われているワイブル分布のｋ＝0.75などを用いるほうが現実的であろう。

参考文献　１）Simiu, E. & Filliben, J.J.: Proc. ASCE, **102**(ST9), 1976, pp.1861-1877.
２）藤野・伊藤・酒井：土木学会論文報告集, **305**, 1981, pp.23-33.
３）合田：港湾技術研究所報告, **27**(1), 1988, pp.31-92.
４）気象庁：観測技術資料, **34**, 1971, 75p.
５）合田・李：土木学会第44回年講, 第Ⅱ部門（水文統計), 1989.