II-73　極値分布関数の棄却基準と年最大日雨量資料への適用について

横浜国立大学工学部　正員　合田良実
日本ＩＢＭ株式会社　李　孝秀

１．まえがき

水文統計を始めとして極値資料の取り扱いでは，データにあてはめるべき分布関数の選択が常に問題となる。一般には，複数の分布関数の中からデータに最も適合すると判断される関数を採択する。しかし，そのための客観的な方法が確立されているとはいいがたい。ここでは逆に，対象とするデータの側からみてあてはめを棄却すべき分布関数の判断基準を検討してみた。あてはめは最小２乗法による場合を対象とし，分布関数として極値Ｉ型分布，ワイブル分布（形状母数４種），および２母数対数正規分布について検討した。

２．標本中の異常値の出現特性

極値資料では標本中に他とかけ離れて大きな値が含まれていることがある。既に角屋[1)]は，標本が対数正規分布あるいは対数極値分布の母集団に帰属するものとして，これから外れる観測値を棄却するための基準を示している。しかし，異常値も他の値と同種類の気象原因で生じたものであるならば，観測値ではなくて不適合の分布関数のほうを棄却すべきであるとも考えられる。その場合は、母関数からランダムに抽出された標本における最大値の分布状況を調べておく必要がある。このため，Ｎ＝10～100 の大きさの標本を母集団から10,000組ずつ疑似一様乱数を使って抽出し，各標本中の最大値x_1を次のように無次元化した。

$$\xi = (x_1 - x_m)/s_x \qquad (1)$$

ただし，x_mはその標本の平均値，s_xは標本の標準偏差である。式（１）の標本最大値の無次元偏差は分布関数の母数の絶対値に影響されない，すなわちパラメーターフリーであり，このことは確認済みである。

式（１）のξの値各10,000個を大きさの順に並べ替え，上から 100番目および 500番目の値を求めることよってξの１％および５％超過限界値が得られたので，これを式（２）のような直線回帰式で近似した。

$$\xi = a + b\,\ln N \qquad (2)$$

各分布関数に対する係数値を求めた結果は表－１のとおりである。なお，対数正規分布については極値データの生の値ではなく，その自然対数を取った値について平均，標準偏差などを計算して処理するものとする。式（２）による近似値とモンテカルロ法による数値実験値との差はおおむね±2%以内であるが，ワイブル分布のｋ＝0.75はξと$\ln N$の関係がやや上向きにカーブ，ｋ＝2.0 や対数正規分布ではやや下向きにカーブしているため，特にＮ＝10などで誤差が大きい傾向がある。

表－１　異常値偏差の係数ａおよびｂの値

分布関数	５％限界値 a	５％限界値 b	１％限界値 a	１％限界値 b
ﾜｲﾌﾞﾙ(k=0.75)	-1.220	1.745	-2.168	2.146
ﾜｲﾌﾞﾙ(k=1.00)	-0.162	1.284	-0.859	1.622
ﾜｲﾌﾞﾙ(k=1.40)	0.673	0.873	0.355	1.092
ﾜｲﾌﾞﾙ(k=2.00)	1.170	0.586	1.165	0.699
極値Ｉ型分布	0.341	0.996	-0.119	1.273
対数正規分布	1.438	0.386	1.585	0.444

３．最小２乗法における相関係数の残差値の分布特性

最小２乗法による分布関数の母数推定の際には，まずプロッティング・ポジション公式として分布関数毎に最も適切なものを使用しないと，再現確率統計量の推定値に偏りを生じる[2)]。データの順序統計量x_iとその基準化変量y_iに対して最小２乗法を適用したときの相関係数ｒはデータに対する分布関数の適合度を最もよく表示する。そこで，前項と同様なモンテカルロ法によって相関係数の残差$\Delta r = 1 - r$の累積分布を求め，これからその５％超過限界値を算出してこれを式(3)のように表示した。

$$\Delta r = A N^B \qquad (3)$$

係数ＡおよびＢの値は分布関数毎に表－２のように算定されている。式（３）のあてはめ誤差は，ワイブル分布のｋ＝2.0 が－3%～＋4%，対数正規分布が－2%～＋3%とやや大きいが，その他は±2%以下である。

また，相関係数の残差の平均値について式（3）のべき乗回帰式を適用して係数を求めた結果も表－2に示している。この場合のあてはめ誤差は５％限界値よりも幾分大きい。特にk＝2.0 のワイブル分布では下方に凸の傾向が強く，±8%程度の誤差を伴う。

表－2　相関係数の残差値の係数AおよびB

分布関数	5％限界値 A	5％限界値 B	平均値 A	平均値 B
ﾜｲﾌﾞﾙ(k=0.75)	0.339	-0.444	0.127	-0.428
ﾜｲﾌﾞﾙ(k=1.00)	0.371	-0.553	0.147	-0.541
ﾜｲﾌﾞﾙ(k=1.40)	0.421	-0.672	0.163	-0.639
ﾜｲﾌﾞﾙ(k=2.00)	0.451	-0.751	0.162	-0.676
極値Ｉ型分布	0.378	-0.614	0.164	-0.616
対数正規分布	0.538	-0.802	0.207	-0.745

4．分布関数の棄却基準

上述の標本中の異常値および最小２乗法適用の際の相関係数の残差値の分布特性を利用すると，データにあてはめるべき分布関数の棄却検定を行なうことが可能である。ここで，標本中の最大値の偏差を用いるものをＤＯＬ(Deviation of OutLier)基準，相関係数の残差値を用いるものをＲＥＣ(REsidue of Correlation coefficient）基準と呼ぶことにする。すなわち、想定した分布関数に対して標本のξ値が表－1の５％限界値を上回る，あるいはΔｒが表－2の５％限界値を上回るときは，データがその分布関数で表される母集団に所属する確率が５％未満であると判断し，その分布関数を棄却する。実データに対する適用事例ではＲＥＣ基準のほうが厳しく，ＤＯＬ基準で棄却されるものは同時に前者でも棄却される場合が多い。

また，最適合の判断基準として著者の一人は先に相関係数が最大のものを採択することを提案した[2)]が、表－2のように相関係数の平均値が分布関数によって異なることを考慮すると，Δｒの平均値に対する比率の最小のものを最適合と判断するほうが妥当ではないかと思われる。この判断基準をここではＭＩＲ(MInimum Ratio of residual correlation coefficient)基準と呼んでおく。

5．年最大日雨量資料に対する適用結果

気象庁観測年報に記載されているもののうち，観測期間ができるだけ長く取れるよう，明治年間からの観測値のある58地点（1987年までの観測年数76～111年）を対象とした。棄却検定の結果を表－3に示す。

表－3　年最大日雨量に対する各種分布関数あてはめの適合度

地点名	a b c d e f	地点名	a b c d e f	地点名	a b c d e f	地点名	a b c d e f	地点名	a b c d e f
網走	▲○◎▲○○	小名浜	▲◎○▲○○	甲府	▲○◎▲○○	京都	◎○▲■▲▲	浜田	◎○▲■▲▲
根室	▲▲○○○◎	宇都宮	▲○○○◎○	飯田	○◎■■■■	大阪	○◎▼■▼▼	下関	○◎■■▼▼
釧路	▲○○■◎○	水戸	○◎○■○○	浜松	○◎○■○▲	神戸	○◎○■▲▲	厳原	▲○○○◎○
札幌	○◎○■○○	銚子	○◎○■○○	名古屋	▲○○○◎○	和歌山	○○○■◎○	福岡	▲○◎○○○
寿都	◎○▲■▲■	八丈島	▲○◎○○○	津	▲○◎▲○○	潮岬	○○○▲◎○	佐賀	○◎■■▲■
函館	○◎▼■○○	横浜	▲○○○◎○	岐阜	▲○◎▲○○	高知	○◎■■▼■	長崎	○◎○▲○○
青森	○◎○■○▲	東京	○◎■■■▼	高山	○◎○■▲▲	徳島	○◎▲■▲○	熊本	◎○▲■▲▲
秋田	▲○○○◎○	熊谷	○◎○▲○○	伏木	○◎○▲○▲	多度津	○○◎▲○○	大分	○◎○■○○
宮古	▲○◎▲○○	前橋	◎▼■■■■	福井	○○○▲◎○	松山	▲○○■◎○	宮崎	○◎■■▲▲
石巻	▲▲○◎○○	新潟	○○◎▲○○	金沢	○○◎■○▲	広島	○◎■■■■	鹿児島	▲▲○◎○○
山形	○◎▲■▲▲	長野	▲○◎○○○	敦賀	▲▲○○○◎	岡山	▲○○▼◎○		
福島	▲○◎▲○○	松本	○○◎▲○▲	彦根	■■■■■■	境	○◎○■○○		

注1）分布関数：　a＝ワイブル（k＝0.75），　b＝ワイブル（k＝1.0），　c＝ワイブル（k＝1.4），
d＝ワイブル（k＝2.0），　e＝極値Ｉ型分布，　f＝対数正規分布．

注2）適合度判定：◎＝ＭＩＲ基準最適合，　○＝棄却基準に該当せず，
▲＝ＲＥＣ棄却基準に該当，　▼＝ＤＯＬ棄却基準に該当，　■＝両棄却基準に該当．

この結果では、指数分布（ワイブルのk＝1.0)の棄却率が約９％と最小であってかつ最適合の割合も高いのに対し，極値Ｉ型分布や対数正規分布は棄却率が約30％もあり，一般性を保ち得ないといえる。

参考文献　1）角屋　睦：農業土木研究，別冊第３号，1962，pp.23-27.
2）合田良実：港湾技術研究所報告，第２７巻第１号，1988，pp.31-92.