

II-106 極値統計における再現期間の信頼区間

横浜国立大学工学部 正員 合田良実
東京都下水道局 関口昌之

1. まえがき

土木工学における極値統計解析の主目的は、与えられた統計資料に基づいて100, 200年等の再現期間に対応する確率統計量を推定し、設計外力選定の目安を与えることである。このための統計解析手法については数多くの研究がなされており、再現確率統計量を求める手順は一通り確立されているといえる。しかし、得られた再現確率統計量がどの程度の標本誤差を伴うものであるかについては、近年ようやく研究成果が発表され始めているに過ぎない（たとえば文献¹⁾）。まして、100年確率などと称する設計外力の再現期間の数値的信頼性についてはほとんど未検討である。ここでは、Monte Carlo 法による数値実験により、標本の統計的変動性に起因する再現期間の信頼区間について検討した結果を報告する。

2. 数値実験の考え方と実験結果

いま一つの極値資料の標本に対して極値分布関数のあてはめを行い、図-1に示すような再現期間 R と極値 x_R の関係を得たとする。再現期間 $\hat{R}$ が指定されたときは、図のA点を求めてその縦軸の値を読み取ることによって $\hat{R}$ 年確率統計量が推定できる。しかし、この標本が母集団を適切に代表している保証はなく、母集団の確率分布は図の破線で示すようなものであるかも知れない。そうすると、極値 x_R の真の再現期間は $\hat{R}$ ではなくて、図のB点に対応する R_0 ということになる。実際のデータについて両者の差を比較することは不可能であるが、数値実験では母集団を固定しておいてこれから十分な数の標本を抽出し、それらの標本に対して極値分布関数をあてはめて $\hat{R}$ と R_0 の統計的差異を調べることは容易である。本研究では、基本的に各分布関数ごとに10,000組の異なる標本を発生させて分布関数のあてはめを行った（最小2乗法を使用）。そして、あらかじめ設定した再現期間 $\hat{R}=10\sim1000$ （年）のレベル毎に真値 R_0 を求めて次のような補助統計量 Q を定義した。

$$Q = \log_{10}(R_0/\hat{R}), \quad R_0 = 10^Q \times \hat{R} \quad (1)$$

母分布関数としては極値I型分布、極値II型分布（形状母数4種）、およびワイブル分布（形状母数4種）の9種類をとりあげ、標本の大きさ（標本中のデータ個数）は $N=10\sim100$ の範囲を対象とした。極値II型分布の関数形としては、合田・小野澤²⁾によるものを用いた。図-2は、極値I型分布の $N=40$ の標本について $R=100$ に相当する Q 値を推定して、その頻度分布を求めた例である。このような頻度分布から90%信頼区間などを推定することは容易である。そこで、本研究では上下5%の範囲、すなわち5%および95%非超過限界値を対象として、その算定のための実験式を試行錯誤によって表-1のように作成した。数値シミュレーション結果に対するこれらの実験式の適合精度は、極値I型分布とワイブル分布では $\pm 4\%$ 以内であり、極値II型分布では $\pm 6\%$ 以内であった。

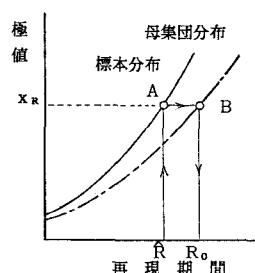
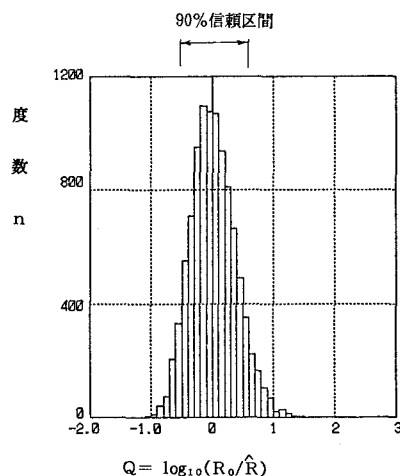


図-1 再現期間の真値と推定値

極値I型分布 [$N=40, \hat{R}=100$]

— 標本数 10,000 組 —

図-2 補助統計量 Q の頻度分布

3. 極値統計資料に対する適用事例

極値資料の中には、数値的に異常に大きなデータを含むものがときどきあり、その場合の取り扱いが問題

となる。異常値の棄却検定問題については先に検討したところであるが³⁾、ここでは大阪の年最大風速、彦根の年最大日雨量、および高知港の年最大潮位について各観測第1位の再現期間について検討する。大阪の1929～1987年の間の観測最大風速は1934年の室戸台風による42.2m/sである。彦根の1895～1987年の期間における観測最大日雨量は1896年9月の596.9mmである。このときの気象状況等は不明であるが、ここでは観測値が信頼できるものとして解析を進めた。高知港の1950～1988年の期間の潮位資料では観測最大潮位が1970年の土佐湾台風によるD.L.上422.9cmである。これらの極値はいずれも第2位以下を大きく引き離しており、各種分布関数のあてはめ結果では極値Ⅱ型分布の形状母数 $k=2.5$ が最適である。そこで、この極値Ⅱ型分布をあてはめたときの観測第1位の再現期間の推定値およびその90%信頼区間を表-1の実験式を用いて求めたところ、表-2の結果が得られた。どのケースも、再現期間の信頼区間は推定値の約1/4～4倍の範囲にまで広がっている。

また、異常に大きなデータを含む極値資料の解析では、第1位のデータを異常値とみなしてこれを除外する場合も少なくない。ただし、数値として疑問があっても、他よりも大きなデータが発生した事実は否定できないと考えられる。そこで最小2乗法適用の際には、プロットング・ポジションを原データの順位のままで計算し、第1位を空欄として第2位以下を対象として母数推定を行った。これについては、慣用的に取り上げられることの多い指数分布をあてはめた結果を表-2に併記した。これをみると、彦根の観測最大日雨量の再現期間は推定値が390万年ときわめて長いものとなり、信頼区間としても1.2万年～32億年と著しく幅が広がってしまう。大阪の風については再現期間の信頼区間が推定値の約1/7～10倍、高知港の潮位では信頼区間の幅が推定値の約1/8～12倍である。分布関数あてはめの検定を行うと、この3ケースとも指数分布のあてはめが棄却されるのであるが、このようなやや無理なあてはめを行うとそれにつれて再現期間の信頼区間の幅が広がることになる。

以上により、極値統計資料に対して極値Ⅰ型分布、極値Ⅱ型分布、およびワイブル分布をあてはめてR年確率統計量を推定する場合について、得られた統計量の再現期間そのものの信頼区間が明らかになった。

参考文献 1) 合田良実：港湾技術研究所報告、第27巻 第1号、1988、pp.31-92。

2) 合田良実・小野澤昌己：極値Ⅱ型分布の特性と推定値の信頼区間（ノート）、土木学会論文集、No.417/Ⅱ-13、1990年5月。

3) 合田良実：極値統計における異常値の棄却基準とその応用、土木学会論文集、No.417/Ⅱ-13。

表-1 再現期間の信頼区間の実験式とその係数

$$\text{極値Ⅰ型分布: } Q = a + b y_R$$

$$\text{極値Ⅱ型分布: } Q = a y_R^{[b + \alpha / (1 + \beta y_R)]}$$

$$\text{ワイブル分布: } Q = a y_R^b$$

分布関数		a	b	α	β	
極値Ⅰ型	95%限界	$0.47N^{-0.63}$	$0.72N^{-0.49}$	—	—	
	5%限界	$-0.12N^{-0.36}$	$-0.55N^{-0.45}$	—	—	
極 値	95%限界	$k=2.5$	$0.73N^{-0.324}$	$0.07N^{0.26}$	0.4	0.05
		$k=3.3$	$0.74N^{-0.338}$	$0.15N^{0.11}$	0.5	0.10
		$k=5.0$	$1.10N^{-0.504}$	$0.17N^{0.18}$	0.5	0.10
		$k=10.0$	$1.04N^{-0.520}$	$0.33N^{0.08}$	0.5	0.10
Ⅱ 型	5%限界	$k=2.5$	$-0.52N^{-0.294}$	0.27	0.4	0.05
		$k=3.3$	$-0.49N^{-0.345}$	0.38	0.5	0.10
		$k=5.0$	$-0.52N^{-0.393}$	0.45	0.5	0.10
		$k=10.0$	$-0.61N^{-0.438}$	0.50	0.5	0.10
ワ イ ブ ル	95%限界	$k=0.75$	$0.79N^{-0.477}$	0.80	—	—
		$k=1.0$	$0.83N^{-0.516}$	1.10	—	—
		$k=1.4$	$0.86N^{-0.533}$	1.55	—	—
		$k=2.0$	$0.89N^{-0.546}$	2.24	—	—
5%限界	$k=0.75$	$-0.58N^{-0.427}$	0.79	—	—	
	$k=1.0$	$-0.59N^{-0.452}$	1.07	—	—	
	$k=1.4$	$-0.59N^{-0.457}$	1.48	—	—	
	$k=2.0$	$-0.57N^{-0.454}$	2.10	—	—	

表-2 極値第1位の再現期間の信頼区間の計算例

項 目	大阪・風 (m/s)	彦根・雨量 (mm)	高知・潮位 (cm)
データ個数 N	59	93	39
第1位極値 x_1	42.2	596.9	422.9
平均値 $\bar{x}$	15.19	100.48	241.8
標準偏差 s	4.934	88.68	38.8
最適の極値Ⅱ型分布 ($k=2.5$) による推定			
尺度母数推定値 A	1.733	19.89	14.20
位置母数推定値 B	13.07	76.19	224.5
再現期間推定値 $\hat{R}$	166.2年	446.8年	111.9年
同上 5%限界値 $R_{5\%}$	39.8年	116.7年	23.8年
同上95%限界値 $R_{95\%}$	719.3年	1871.年	521.8年
第1位を除外した指数分布による推定			
尺度母数推定値 A	3.860	35.26	31.20
位置母数推定値 B	11.12	61.45	203.2
再現期間推定値 $\hat{R}$	3140.1年	3,93E+6年	1142.8年
同上 5%限界値 $R_{5\%}$	423.1年	1.23E+4年	140.8年
同上95%限界値 $R_{95\%}$	3.1E+4年	3.17E+9年	13520.年