

最小2乗法に基づく極値統計解析システムの精度の検討

山口正隆*・大木泰憲**・前川隆海***

1. 緒 言

種々の極値統計解析システムの中で、合田(1988, 1990)により提案された「最小2乗法に基づく極値統計解析システム」は、長期間の資料が取得されていない波高の極値統計解析の分野において、年最大値資料および極大値資料の両者に適用できる有用な方法である。このため、わが国における沿岸構造物の設計波の選定に広範に利用されているが、このシステムの精度についての詳しい検討はあまり行われていない。

そこで本研究では、両極値資料を対象として、広範囲に入力条件を設定したモンテカルロシミュレーション結果より得られた波高資料に対する極値統計解析結果に基づいて、確率波高およびその標準偏差の推定精度に及ぼす候補確率分布数、年平均発生数、データ採択率、確率分布形状および未知母分布の影響を詳細に検討する。

2. システムの精度の検討方法

(1) システムの概要

本システムは、Gumbel分布と形状母数を $k=0.75$, 1.0, 1.4, 2.0とする4種類の3母数Weibull分布の合計5種類、あるいはGumbel分布と形状母数を $k=2.5$, 3.33, 5.0, 10.0とする4種類のFT-II型分布の合計5種類(いずれもkind5と呼ぶ)あるいは合計9種類(kind9)の候補確率分布を極値資料にあてはめて、母数を最小2乗法で決定したのち、相関係数あるいはMIR基準(本研究では、相関係数基準を使用)に従って、候補確率分布の中から最適確率分布を選択するとともに、確率波高およびその標準偏差を推定するものであり、最小2乗法に特有のプロットングポジション公式が工夫されている。また、本システムは事象の年平均発生数 $\lambda=N_T/K$ (N_T : 極値資料総数, K : 年数)およびデータ採択率 $\nu=N/N_T$ (N : 採択資料数)を導入することによって、年最大値資料のみならず極大値資料にも適用可能であるし、候補確率分布を棄却するための2つの基準(DOL基準、

REC基準)を含む。

合田により用いられたGumbel分布、Weibull分布およびFT-II型分布の未超過確率分布 $F(x)$ はつぎの形で表される。

$$\left. \begin{aligned} F(x) &= \exp[-\exp\{-(x-B)/A\}]; & -\infty < x < \infty \\ F(x) &= 1 - \exp[-\{(x-B)/A\}^k]; & B \leq x < \infty \\ F(x) &= \exp[-\{1+(x-B)/kA\}^{-k}]; & B-KA \leq x < \infty \end{aligned} \right\} \dots\dots\dots (1)$$

ここに、 x : 確率変数, k : 形状母数, A : 尺度母数, B : 位置母数, である。

(2) モンテカルロシミュレーションの方法

母集団の確率分布が明らかな場合の本システムの精度を検討するため、モンテカルロシミュレーションを実施する。Gumbel分布、Weibull分布およびFT-II型分布のinverse formはそれぞれ、

$$\left. \begin{aligned} x &= B - A \cdot \log\{-\log F(x)\} \\ x &= A[-\log\{1-F(x)\}]^{1/k} + B \\ x &= B + kA[(-\log F(x))^{-1/k} - 1] \end{aligned} \right\} \dots\dots\dots (2)$$

で表されるので、0~1間の一様乱数により未超過確率 F を与えれば、上記の確率分布に従う波高資料を作成することができる。

シミュレーションの順序は次のようである。

① 母集団の確率分布と母数および年平均発生数 λ とデータ採択率 ν を指定し、確率波高($R=50$ 年, 100年, 200年, 500年, 1000年の5ケース)の真値を式(2)において $F=1-1/\lambda R$ として計算する。

② 式(2)に一様乱数を与えて、母分布から対象年数 K の年平均発生数 λ 倍の標本($N_T=\lambda K$, 最大24000)を発生させ、そのうち上位 $N(=\nu N_T)$ 個の標本を抽出する。年最大値資料を対象とする場合、 $N \leq K=N_T$ である。

③ 発生標本から最小2乗法により候補確率分布の母数を推定するとともに、相関係数 ρ 、および最適確率分布についての確率波高 H_R と合田式による標準偏差 σ_R を求める。

④ 対象年数 $K=10, 20, 30, 40, 50, 70, 100, 200, 500, 1000$ 年の10ケースの場合に、②と③を M 回($M=5000$)繰り返し、相関係数 ρ の平均値、確率波高の平均値とbias ΔH (確率波高の平均値-確率波高の真値)、標

* 正 会 員 工 博 愛媛大学教授 工学部環境建設工学科

** 正 会 員 工 修 東亜建設工業株式会社

*** 学生会員 愛媛大学大学院 理工学研究科

標準偏差 $Var^{1/2}$, 合田式による確率波高の標準偏差の平均値などの各種誤差統計量を算定する。

一方, 未知母分布の影響を考える場合には, 100 年確率波高を 12 m, 1000 年確率波高を 15 m, 形状母数を既知として尺度母数および位置母数を求めた 3 母数 Weibull 分布, GEV 分布, 3 母数対数正規分布, 3 母数 Gamma 分布, 3 母数対数 Gamma 分布, 3 母数超 Gamma 分布, および Gumbel 分布 (2 母数分布) の 7 種類の確率分布に基づいて, 各確率変量を 1000 回ずつ均等に発生させることにより, 未知母分布という条件を近似し, 同様の解析を行う。また, 比較のために, 母分布を Gumbell 分布, Weibull 分布および FT-II 型分布に固定した場合 ($M=7000$) にもシミュレーションを行う。

(3) シミュレーションの入力条件

シミュレーションは, 極大値資料 ($\lambda > 1, \nu \leq 1$) および年最大値資料 ($\lambda = 1, \nu \leq 1$) の場合に, 本システムの精度 (確率波高の bias および合田式による標準偏差) に及ぼす資料年数 K , 候補確率分布数, 年平均発生数 λ , データ採択率 ν , 形状母数 k および未知母分布の影響を検討することを目的として実施するので, 入力条件をつぎのように設定する。すなわち, Weibull 分布の形状母数は $k=0.75, 0.85, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.5$ の 9 ケース ($A=4.0, B=1.0$ に固定), Gumbel 分布の母数は $A=1/0.72, B=4.50$, FT-II 型分布の形状母数は $k=2.5, 2.86, 3.33, 4.0, 5.0, 6.67, 10.0, 15.0$ ($A=1.0, B=5.0$) の 8 ケース, データ採択率は $\nu=1/8, 2/8, 3/8, 4/8, 6/8, 1$ の 6 ケース (年最大値資料の場合, 資料数が少ないため, $K=10$ に対して $\nu=1/8$ の場合を省略), 極大値資料における年平均発生数は $\lambda=8$ (λ の影響を調べる場合には, $\lambda=16, 24$ も使用), 候補確率分布の数は 5 あるいは 9, である。

3. シミュレーション結果の検討

(1) 候補確率分布数の影響

図-1 は母分布を Weibull 分布 (形状母数 $k=1.4$) とした $\lambda=\nu=1$ の年最大値資料について, 候補確率分布数 (kind 5 あるいは kind 9) の誤差統計量に及ぼす影響を検討した結果であり, 図は上から相関係数 ρ , 100 年確率波高の真値で無次元化した, 100 年確率波高の推定値の bias $\Delta \bar{H}$, 標準偏差 $\bar{Var}^{1/2}$, および 100 年確率波高の $Var^{1/2}$ で無次元化した合田式による 100 年確率波高の標準偏差 (既知母分布の場合) の平均値 $\bar{\sigma}$ をそれぞれ % 表示したものと対象年数 K との関係を示す。

この図によれば, 対象年数 (資料数) の増加とともに, 相関係数は増加して 1 に漸近し, bias および $Var^{1/2}$ は単調に 0 に近づくという意味で合理的な結果が得られている。また, $Var^{1/2}$ の減少の度合は bias より緩やかである

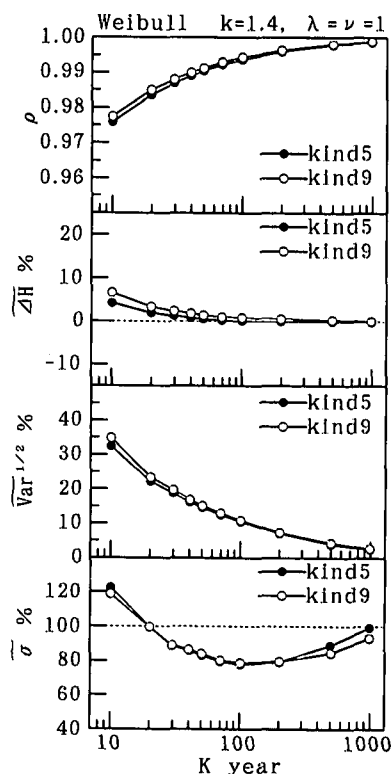


図-1 誤差統計量に及ぼす候補確率分布数の影響

ことから, 確率波高の精度を向上させるためには, 対象年数を増加させる必要がある。一方, 標準偏差算定式は $\pm 20\%$ の誤差内の値を与えるが, 資料年数の増加とともに, 単調に $Var^{1/2}$ に漸近する挙動を示さない。

候補確率分布数の影響をみると, 相関係数は kind 9の方が大きい。候補確率分布数が多いほど, 適合度の高い分布が抽出されるので, 相関係数が大きいのは当然である。しかし, bias は, kind 5 の場合より, kind 9 の場合に大きい。一般に, Weibull 分布および Gumbel 分布に対する bias の値は kind 9 の場合が kind 5 の場合より大きく, FT-II 型分布の場合には逆になる。この傾向は $\nu < 1$ の場合も同様である。また, bias は形状母数により正負の値をとるので, いずれの候補確率分布数の場合の bias がより 0 に近いかは, Weibull 分布あるいは FT-II 型分布の形状母数によって変わる。たとえば, $k=1.4$ の Weibull 分布では kind 5 の場合の bias が, 一方, $k=0.75$ では kind 9 の場合の bias が 0 に近い。ただし, 全般的には kind 5 の場合の bias が 0 に近い傾向を示す。

$Var^{1/2}$ はいずれの極値資料でも kind 5 の場合より kind 9 の場合に大きい。これは, 候補分布数が多い場合, 各ケースの資料に対する適合度が向上するため, 全ケースに対する変動は大きくなることを意味する。ただし,

FT-II型分布資料に対する最適確率分布はFT-II型分布と判定されるケースが多い(分布が裾を長くひくため)であろうから、 $Var^{1/2}$ は候補確率分布数にほとんど依存しない。また、標準偏差算定式の精度も候補確率分布数にあまり依存しないが、kind 5の場合にやや良好な傾向を示す。

以上をまとめると、適合度の観点からは候補分布数としてkind 9が適切であるが、確率波高の推定精度(biasおよび標準偏差)の観点からは、kind 5が適切である。極値統計解析の立場に立てば、biasおよび標準偏差が小さい方が好ましいので、本研究では、候補確率分布数として、kind 5を採用する。

なお、候補確率分布にかかわらず、①相関係数 ρ は、対象年数、年平均発生数、データ採択率(いずれも採択資料数に相当)、候補確率分布数、形状母数が大きいほど、1に近い値をとる、②標準偏差 $Var^{1/2}$ は対象年数、年平均発生数、データ採択率、形状母数が大きいほど小さく、候補確率分布数が多いほど大きくなる、という特性が共通して見出される。

(2) 年平均発生数の影響

シミュレーションでは、同一対象年数に対して、年平均発生数 λ が大きいほど、採択資料数も増える。図-2は母分布をWeibull分布($k=2.0$)として λ を3種類($\lambda=8, 16, 24$)変えた極大値資料($\nu=0.375$)に対する誤差統計量と対象年数の関係を示す。これから、①biasは λ が大きいほど小さい正値をとり、対象年数の増加とともにほぼ消失するが、図-2のように、 $\nu \leq 0.375$ では、年

平均発生数の影響はあまりみられない、② $Var^{1/2}$ も λ および K の増加とともに減少し、0に漸近する、③標準偏差算定式は $K \leq 200$ では λ が大きいほど $Var^{1/2}$ を過小評価する、という傾向が見出される。

FT-II型分布の場合、biasおよび $Var^{1/2}$ に及ぼす λ の影響は小さい。また、標準偏差算定式は λ が大きいほど標準偏差を過小評価するが、その度合はWeibull分布の場合より著しい。

(3) データ採択率の影響

図-3は $k=1.4$ のWeibull分布に対応する極大値資料($\lambda=8$)の場合に、誤差統計量に及ぼすデータ採択率の影響を示したものである。biasは形状母数によって正負をとりうるが、採択率が大きいほど、その絶対値は小さくなり、確率波高の推定精度が向上する。これは、Weibull分布およびFT-II型分布の場合のいずれの形状母数についてもいえる。しかし、Gumbel分布の場合には逆の傾向を示す。 $Var^{1/2}$ はすべての場合について ν が大きいほど小さい値をとる。また、標準偏差算定式による値と $Var^{1/2}$ の差は、データ採択率が大きくなるほど広がり、 $\nu=0.125$ の場合を除いて、最大40%(FT-II型分布では、60%)の過小評価傾向を生じる。

一方、年最大値資料の場合には、いずれの分布($k=2.0$ のWeibull分布を除く)についても、biasおよび $Var^{1/2}$ に及ぼす ν の影響は全般的に小さい。また、標準偏差算定式は同一対象年数に対して ν が小さいほど大きい値を与え、 ν が大きい場合には過小評価傾向、 ν が小さい場合には過大評価傾向を示す。ただし、通常の極値統計解

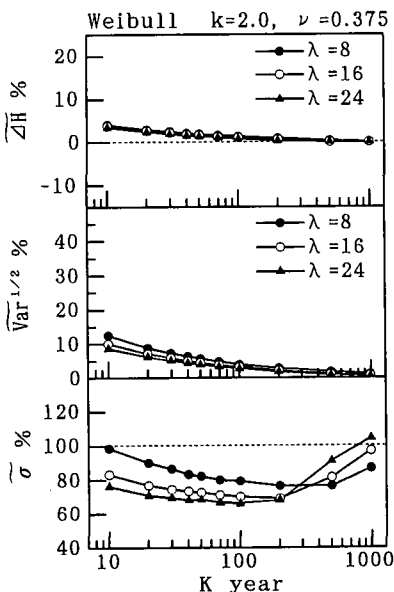


図-2 誤差統計量に及ぼす年平均発生数の影響

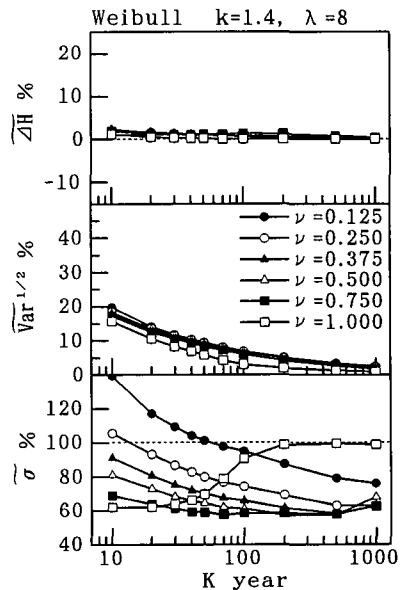


図-3 誤差統計量に及ぼすデータ採択率の影響

析で用いられる $K \leq 100$ で $\nu=1$ の年最大値資料の場合には、その最大誤差は $\pm 25\%$ 程度である。

結局、両極値資料いずれの場合にも、データ採択率が高いほど、採択資料数が増大するので、適合度は向上し、標準偏差も減少する。ただし、データ採択率の影響は、年最大値資料の場合には、あまり現れず、極大値資料の場合に、相対的に大きい。また、標準偏差算定式は極大値資料の場合、データ採択率が高いほど過小評価傾向を生じるし、年最大値資料の場合、データ採択率の大小および資料年数に依存して過小・過大評価あるいは適正評価というように、その精度が変わる。

(4) 形状母数の影響

図-4 は Weibull 分布および FT-II 型分布に対する誤差統計量 (いずれも kind 5) に及ぼす形状母数の影響を $\lambda=8$, $\nu=0.375$ の極大値資料について示したものである。Weibull 分布の場合、 k が小さいほど負の有意な bias を、 k が大きいほど正の有意な bias を生じるが、 $k=1.0 \sim 1.4$ では bias はほとんど生じない。種々の気象海象要素を対象とした現実の極値統計解析では、年最高気

温のように年変化の小さい極値資料の分布形状は最小 2 乗法で用いられる最も尖鋭な $k=2.0$ の Weibull 分布より尖っていることがあるので、現行のシステムはこの場合の再現確率値を有意な水準で過大評価することになる。一方、FT-II 型分布における形状母数の影響も Weibull 分布と同様であり、 $k=5.0 \sim 15.0$ における bias は無視しうるが、 $k \leq 3.33$ では負の有意な bias を生じる。とくに、最も裾の長い分布である $k=2.5$ の場合には、bias は 10% 以上に達し、しかも対象年数が増加しても必ずしも減少しない傾向にある。また、標準偏差算定式はいずれの分布の場合にも $Var^{1/2}$ を過小評価する傾向 ($K < 100$ に対して最大 35% 程度) にあるが、形状母数による変化は一貫しない。

図-5 は Weibull 分布および FT-II 型分布より生成した年最大値資料 ($\lambda=\nu=1$) について形状母数の影響を検討した結果である。これらは極大値資料の場合と同様の特性を示すが、bias が極大値資料の場合より上側に図示される。このため、極大値資料に対する結果と比べて、Weibull 分布の場合には $k \geq 1.4$ 、FT-II 型分布の場合に

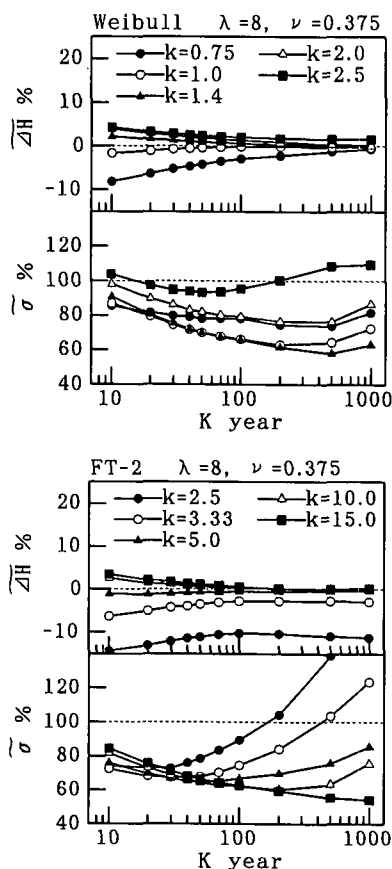


図-4 誤差統計量に及ぼす形状母数の影響 (1)

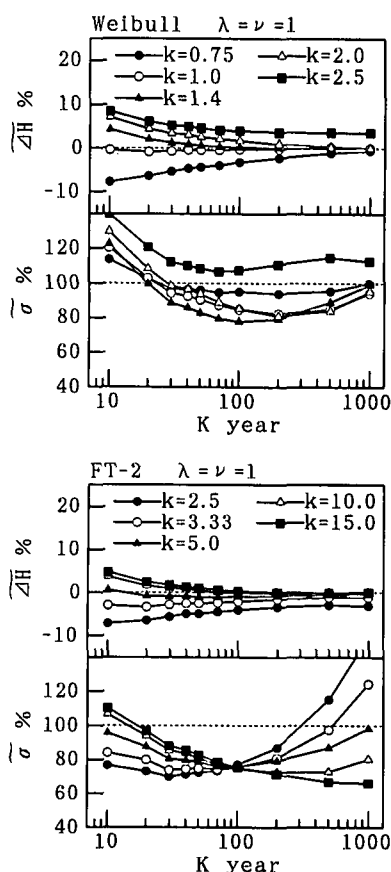


図-5 誤差統計量に及ぼす形状母数の影響 (2)

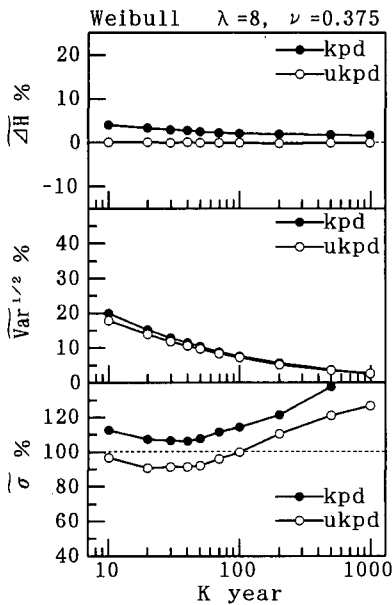


図-6 誤差統計量に及ぼす未知母分布の影響

は $k \geq 10.0$ で bias が増加し、確率波高の過大評価傾向が若干助長されるが、それぞれ $k \leq 1.0$ および $k \leq 3.33$ では bias の絶対値が減少し、確率波高の過小評価傾向が改善される。これは $k=2.5$ の FT-II 型分布の場合に著しい。一方、標準偏差算定式の挙動は全く異なっている。Weibull 分布 ($k=2.5$ を除く) より生成した年最大値資料の場合、 $K=20$ 付近で 100% 線を正から負へ横切って、 $K=1000$ では 100% 線に近づき、誤差も最大 $\pm 20\%$ 程度である。また、FT-II 型分布の場合、形状母数の値によって精度は異なり、 $K < 100$ では形状母数の値が小さいほど過小評価傾向が顕著になる。

(5) 未知母分布の影響

図-6 は Weibull 分布 (既知母分布, kpd) および 7 種類の確率分布 (未知母分布, $ukpd$) より生成した極大値資料 ($\lambda=8$) を対象として、極値統計解析に kind 5 の Weibull 分布を使用した場合における未知母分布の影響を $\nu=0.375$ の場合に示したものであり、図中には既知母分布および未知母分布の場合の標準偏差算定式による結果がそれぞれ与えられる。

既知母分布条件および未知母分布条件に対する結果を比べると、一般に、前者の相関係数が大きく、bias およ

び $Var^{1/2}$ が小さいという傾向が期待される。しかし、本シミュレーションケースでは、Weibull 分布 (両極値資料) の場合 bias および $Var^{1/2}$ 、FT-II 型分布 (両極値資料) の場合相関係数、Gumbel 分布 (極大値資料) の場合 bias に関して、逆の傾向が見出される。また、標準偏差算定式の精度はとくに未知母分布の場合にかなり良好であるが、年最大値資料 ($\lambda=\nu=1$) の場合には、 $K=10$ で約 60% の過大評価から $K=100$ で誤差 0% に激減し、 $K=1000$ で再び 20% の過大評価に転じる。なお、未知母分布の場合の標準偏差算定式は FT-II 型分布に対しては提示されていない。

4. 結 語

本研究で得られた結果はつぎのように要約される。

① 候補確率分布の数が多いと、適合度は向上するが、標準偏差は増大する。しかし、bias の挙動は入力確率分布により異なる。

② 年平均発生数 λ および対象年数 K の増加とともに、適合度は向上し、bias および標準偏差は減少する。

③ データ採択率 ν が高いほど、適合度は向上し、標準偏差は減少するが、bias については必ずしもこれがない。また、データ採択率の影響は年最大値資料より極大値資料において大きい。

④ Weibull 分布に基づく極値統計解析システムの場合、形状母数 k が小さいほど負の有意な bias を、 k が大きいほど正の bias を生じる。この傾向は FT-II 型分布に基づくシステムの場合も同様である。

⑤ 未知母分布の場合の bias や標準偏差は既知母分布の場合より必ずしも大きいとはいえず、資料分布にあてはめる確率分布の種類、あるいは誤差統計量の種類によって相対的大小関係は変化する。

⑥ 合田による標準偏差算定式の誤差は、年最大値資料の場合には最大 $\pm 20 \sim 25\%$ 程度であるが、極大値資料の場合には、データ採択率の増加とともに過小評価傾向が助長され、誤差が増大する。

参 考 文 献

- 合田良実 (1988): 極値統計におけるプロットングポジション公式ならびに推定値の信頼区間に関する数値的検討, 港湾技術研究所報告, 第 27 巻第 1 号, pp. 31-92.
- 合田良實 (1990): 港湾構造物の耐波設計一波浪工学への序説一, 鹿島出版会, 333 p.