

極値統計解析における順序統計量の分散を考慮した母数推定法

泉 宮 尊 司*・斎 藤 雅 弘**

1. 緒 言

定期的な波浪観測の目的の1つには、波浪の観測データにより極値統計解析を行い、確率波高を精度よく推定することにある。しかしながら、定期的な波浪観測が開始されてから20年余りしか経っておらず、数少ないデータを用いて確率波高を推定しなければならない。このため、極値統計解析において推定精度の高い母数推定法や信頼区間の推定法の研究が望まれている。

これまで、極値統計解析における母数推定法には、積率法、最小二乗法、最尤法および確率加重平均積率法などが用いられてきた(合田, 1988, 山口ら 1995)。これらの中で最尤法は、統計学者により数多く研究され、理論的な研究が最も進んでいるが、母数を推定する際に、一般的には非線形方程式を解かなければならないこと、および標本数が少ない場合には、不偏性の条件を満たさないなどの欠点があり、実務者が設計外力を決定する際に用いるにはやや難しい側面もある。

一方、積率法や最小二乗法は、その取扱いが簡単であるために実務レベルにおいて、よく用いられてきたが、推定された確率統計量の分散、すなわち推定誤差がやや大きいという欠点を有している。この推定誤差が大きくなる理由の1つに、推定される母数が極値資料の最大値の変動に大きく左右されやすいことが挙げられる。そこで本研究では、従来の最小二乗法を改良するために、各順序統計量およびその分布関数の分布が漸近的にガウス分布に従うことを利用して、各順序統計量の分散を漸近理論により評価し、それらの分散を考慮した母数推定法を提案することを研究の目的とする。

2. 順序統計量の漸近分布と分散

極値統計解析におけるこれまでの最小二乗法では、適切なプロットング公式を用いて非超過確率を算定し、仮定した分布関数より基準化変数を求めて、極値データとの残差を最小にするように母数を求めている。しかしながら、プロットング公式を用いた段階で、極値資料

に順位を付けているために、それぞれの分散はもはや母集団の分散とは異なり、順序統計量の分散を用いなければならない。各順序統計量の分散がほぼ等しければ、特に問題はないが、一般的には大きい方からの順位が1位、すなわち最大値の分散は他の順序統計量の分散よりもかなり大きく、中位程度の順序統計量の分散の数倍にもなることがある。このことから、従来の最小二乗法では分散を同一とみなしているのが、相対的に最大値の影響が大きめにでていると言える。そこで本研究では、まず順序統計量の分散を簡便に推定できる方法を考えることにする。

ここで、極値資料の分布関数を $F(x)$ 、その確率密度関数を $f(x)$ とし、 N 個の標本のうち小さい方から昇順に番号を付した順序統計量を x_i とする。この時、 i 番目の順序統計量 x_i の確率密度関数 $p(x_i)$ は、次式で与えられる(Rao, 1977)。

$$p(x_i) = \frac{N!}{(i-1)!(N-i)!} \cdot [F(x_i)]^{i-1} [1-F(x_i)]^{N-i} f(x_i) \quad \cdots (1)$$

順序統計量 x_i に対して割り当てられる非超過確率 $\xi_i = F(x_i)$ の確率密度関数 $\phi(\xi_i)$ は、確率変数の変換により次式で与えられる。

$$\phi(\xi_i) = \frac{N!}{(i-1)!(N-i)!} (\xi_i)^{i-1} (1-\xi_i)^{N-i} \quad \cdots (2)$$

ここで、順序統計量 x_i の漸近分布を直接的に求めるのはかなり複雑な計算をしなければならないので、非超過確率 $\xi_i = F(x_i)$ の漸近分布を求めた後、Rao (1977) による収束定理を利用して x_i の分散を求めることにする。いま、順序統計量 x_i に対応する非超過確率の最適値 p を、

$$p = \frac{i-\alpha}{N+\beta} \quad \cdots (3)$$

と表す。ここに、 i は昇順位数、 N は標本数、 α および β は、1 よりも小さいある正の実数とする。ここで、

$$q = 1-p = \frac{N-i+\alpha+\beta}{N+\beta} \quad \cdots (4)$$

とおき、次の変数変換を行うことにする。

$$z_i = \frac{\sqrt{N}(\xi_i - p)}{\sqrt{pq}} \quad \cdots (5)$$

* 正会員 工博 新潟大学教授 工学部建設学科

** 工修 新潟県

この変換により、 z_i の確率密度関数 $\psi(z_i)$ は、

$$\psi(z_i) = \frac{1}{\sqrt{N}} \frac{N!}{(i-1)!(N-i)!} p^{i-1/2} q^{N-i+1/2} \cdot \left(1 + z_i \sqrt{\frac{q}{Np}}\right)^{i-1} \left(1 - z_i \sqrt{\frac{p}{Nq}}\right)^{N-i} \dots\dots\dots (6)$$

与えられる。上式の自然対数をとり、式(3)および式(4)とスターリングの公式を用いると、次の関係式を得る。

$$\begin{aligned} \log \psi(z_i) &= \log \left\{ \frac{1}{\sqrt{2\pi}} \frac{1}{e} \left(1 - \frac{q}{Np}\right)^{pN-q+1/2} \left(1 - \frac{p}{Nq}\right)^{qN-p+1/2} \right\} \\ &\quad + \frac{1}{\sqrt{N}} \left(\sqrt{\frac{p}{q}} \sqrt{\frac{p}{q}} - \sqrt{\frac{q}{p}} \sqrt{\frac{q}{p}} \right) z_i - \frac{1}{2} \left(1 - \frac{q^2}{Np} - \frac{p^2}{Nq}\right) z_i^2 \\ &\quad \dots\dots\dots (7) \end{aligned}$$

ここで、 $N \rightarrow \infty$ とすると、上式は

$$\log \psi(z_i) = \log \left(\frac{1}{\sqrt{2\pi}} \right) - z_i^2/2 \dots\dots\dots (8)$$

となる。したがって、 z_i は漸近的に平均値が0で分散が1の正規分布に近づくことを示している。平均値が μ 、分散が σ^2 である正規密度関数を $N(\mu, \sigma^2)$ と書くことにすると、上式の関係により $\sqrt{N}(\xi - p)$ の漸近分布は $N(0, pq)$ となる。

ところで、非超過確率 ξ_i と順序統計量 x_i との間には、

$$\xi_i = F(x_i) \dots\dots\dots (9)$$

なる関係があるので、 $\xi_i = p$ におけるヤコビアンは、

$$\left(\frac{dx_i}{d\xi_i} \right)_{\xi_i=p} = \frac{1}{f(x_{ip})} \dots\dots\dots (10)$$

となる。ここに、 x_{ip} は非超過確率 ξ_i が、 $\xi_i = p$ となる時の x_i の値で、 $p = F(x_{ip})$ なる関係より求められるものである。Rao (1977) の収束定理より、漸近的に

$$\sqrt{N}(x_i - x_{ip}) \rightarrow N\left(0, \frac{pq}{[f(x_{ip})]^2}\right) \dots\dots\dots (11)$$

$$(x_i - x_{ip}) \rightarrow N\left(0, \frac{pq}{N[f(x_{ip})]^2}\right) \dots\dots\dots (12)$$

となることが分かる。したがって、順序統計量 x_i の分散 $\sigma^2(x_i)$ は、上の関係により漸近的には次式で評価される。

$$\sigma^2(x_i) = \frac{pq}{N[f(x_{ip})]^2} \dots\dots\dots (13)$$

式(13)の $f(x_{ip})$ の値は、FT-I型分布(グンベル分布)の場合には、

$$\frac{i-\alpha}{N+\beta} = \exp\left\{-\exp\left[-\frac{x_{ip}-B}{A}\right]\right\} \dots\dots\dots (14)$$

なる関係が成立するために、

$$f(x_{ip}) = \frac{1}{A} \frac{i-\alpha}{N+\beta} \left[-\ln\left(\frac{i-\alpha}{N+\beta}\right)\right] \dots\dots\dots (15)$$

となる。ここに、 A は尺度母数、 B は位置母数である。

したがって、FT-I型分布の場合、式(13)より x_i の分散は、

$$\sigma^2(x_i) = \frac{A^2(N-i+\alpha+\beta)}{N(i-\alpha)\left[\ln\left(\frac{i-\alpha}{N+\beta}\right)\right]^2} \dots\dots\dots (16)$$

ワイブル分布の場合も同様な計算を行うと、

$$\begin{aligned} \sigma^2(x_i) &= \frac{A^2(i-\alpha)}{Nk^2(N-i+\alpha+\beta)\left[-\ln\left(\frac{N-i+\alpha+\beta}{N+\beta}\right)\right]^{2(k-i)/k}} \\ &\quad \dots\dots\dots (17) \end{aligned}$$

を得る。ここに、 k は形状母数である。式(16)および式(17)より、順序統計量 x_i の分散は、尺度母数 A の2乗に比例し、漸近理論により簡単に評価できる。

3. 分散を考慮した母数推定法

これまでの最小二乗法では、極値資料を順番に並べて、それぞれに対応した非超過確率を割り当て、確率分布関数に応じて基準化変量 y_i の値を求めて、次の量を最小にするように母数を決定している。

$$\sum_{i=1}^N (x_i - Ay_i - B)^2 \rightarrow \min \dots\dots\dots (18)$$

上式は、それぞれの残差の分散が等しい場合には特に問題はないが、極値資料の最大値のように大きな分散を持つ場合には、その影響が大きめにでることが予想される。そこで、順序統計量の分散の大きさを考慮して、

$$\sum_{i=1}^N \frac{(x_i - Ay_i - B)^2}{\sigma^2(x_i)} \rightarrow \min \dots\dots\dots (19)$$

とすることを考える。この時、FT-I型分布の場合には、次式を最小化することになる。

$$\begin{aligned} \sum_{i=1}^N \frac{(x_i - Ay_i - B)^2}{\sigma^2(x_i)} &= \frac{N}{A^2} \sum_{i=1}^N \frac{(i-\alpha)\left[\ln\left(\frac{i-\alpha}{N+\beta}\right)\right]^2}{(N-i+\alpha+\beta)} (x_i - Ay_i - B)^2 \\ &= \frac{N}{A^2} \sum_{i=1}^N w_i^2 (x_i - Ay_i - B)^2 \dots\dots\dots (20) \end{aligned}$$

ここに、 w_i^2 は次式で定義される重み係数である。

$$w_i^2 = \frac{(i-\alpha)\left[\ln\left(\frac{i-\alpha}{N+\beta}\right)\right]^2}{(N-i+\alpha+\beta)} \dots\dots\dots (21)$$

式(20)の値を最小とする母数 A および B を求める式は、非線形となるために計算がやや複雑となる。これを避けるために、次のように考えることにする。いま、母集団の分布関数が指定されて標本が N 個得られたとする。 A^2 の値は、母集団の分散に比例した量であり、この分散で規格化した誤差の総和を最小にすることとする。したがって、 $\sum_{i=1}^N w_i^2 (x_i - Ay_i - B)^2$ の値を最小化する母

数 A および B を求めることにする。この時、それらは次の線形式より求められる。

$$A \sum_{i=1}^N w_i^2 y_i + B \sum_{i=1}^N w_i^2 y_i = \sum_{i=1}^N w_i^2 x_i y_i \dots\dots\dots (22)$$

$$A \sum_{i=1}^N w_i^2 y_i + B \sum_{i=1}^N w_i^2 = \sum_{i=1}^N w_i^2 x_i \dots\dots\dots (23)$$

上式において、 $w_i=1, (i=1, N)$ とすると、従来の最小二乗法の関係式となる。また、式 (22) の右辺に $(1/A) \cdot \sum_{i=1}^N w_i^2 (x_i - A y_i - B)^2$ を付加すると、非線形関係式が得られる。

4. プロットング公式の不偏性条件式

プロットング公式を用いる場合、不偏性の条件を満たす公式を用いるのが望ましい。これまでに、様々な議論があり(たとえば、合田, 1989)、数値シュミレーションを行って確認するしかなかった。しかしながら、最小二乗法の場合($w_i=1$)、式 (22) および式 (23) を解くことによって、推定される母数 $\hat{A}$ および $\hat{B}$ は、順序統計量 x_i の線形結合式で与えられる。

$$\hat{A} = \sum_{i=1}^N g_i x_i, \quad \hat{B} = \sum_{i=1}^N h_i x_i \dots\dots\dots (24)$$

ここに、 g_i および h_i は次式で表される。

$$g_i = k_i - \frac{1}{N} \sum_{j=1}^N k_j, \quad h_i = \frac{1}{N} - \bar{y} k_i - \frac{\bar{y}}{N} \sum_{j=1}^N k_j \dots\dots (25)$$

$\bar{y}$ および k_i は、基準化変量 y_i を用いて、

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad k_i = \frac{y_i - \bar{y}}{\sum_{j=1}^N (y_j - \bar{y})^2} \dots\dots\dots (26)$$

と表される。式 (24) において期待値をとると、バイアスがない場合には真の母数と等しくなることから、

$$E[\hat{A}] = \sum_{i=1}^N g_i E[x_i] = A \dots\dots\dots (27)$$

$$E[\hat{B}] = \sum_{i=1}^N h_i E[x_i] = B \dots\dots\dots (28)$$

となる。詳細な計算は省略するが、不偏性の条件式は、

$$\sum_{i=1}^N g_i I_i = 1, \quad \sum_{i=1}^N h_i I_i = 0 \dots\dots\dots (29)$$

となる。ここに、 I_i は FT-I 型分布の場合、

$$I_i = \frac{N!}{(i-1)!(N-i)!} \int_0^1 [-\ln(-\ln t)] t^{i-1} \cdot [1-t]^{N-i} dt \dots\dots\dots (30)$$

となる。式 (29) の g_i および h_i は、プロットング公式における α および β の関数となるために、これを解くことによって、バイアスなしの公式が得られる。

5. 数値シュミレーション結果とその考察

本研究に用いた母集団の確率分布関数は、FT-I 型分布およびワイブル分布である。まず、確率分布関数と母

数を設定し、一様乱数を発生させて N 個の標本を作成する。これを 5000 回繰り返して、5000 個の資料に対して順序統計量の分散を求めると同時に、従来の母数推定法および順序統計量の分散を考慮した推定法により、30 年、50 年および 100 年等の再現確率統計量のバイアスおよび標準偏差を算定した。なお、標本数 N は、 $N=10, 20, 30, 40, 50, 60, 100$ を含む最大で 11 通り設定して行っている。

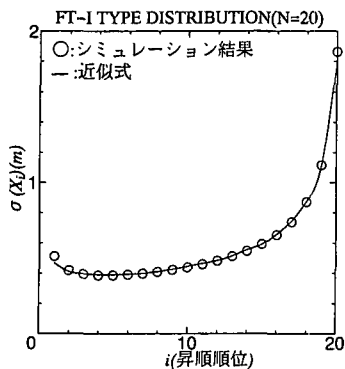
図-1 および図-2 は、FT-I 型分布 ($A=1.39, B=4.5$) の順序統計量 x_i の標準偏差と昇順順位との関係を示したものである。図-1 は標本数が $N=20$ の時の結果であり、図-2 は標本数が $N=50$ の時のものである。図中の○印は、数値シュミレーションにより得られた結果で、実曲線は漸近理論より得られた式 (16) によって計算されたものである。ここでは、プロットング公式として Gringorten (1963) の公式を用いている。標本数が $N=20$ と比較的小さい場合においても、漸近理論値との一致度は良好であることが分かる。

図-3 に、ワイブル分布 ($A=4.76, B=0.39, k=2.0$) の順序統計量 x_i の標準偏差と昇順順位との関係を示したものである。プロットング公式は、合田 (1988) による修正 Petruaskas and Aagaard 公式を用いている。形状母数が $k=2.0$ の時には、順序統計量の値が大きいくほど、その分散が大きくなっているが、数値シュミレーション結果と理論値は良い一致を示している。

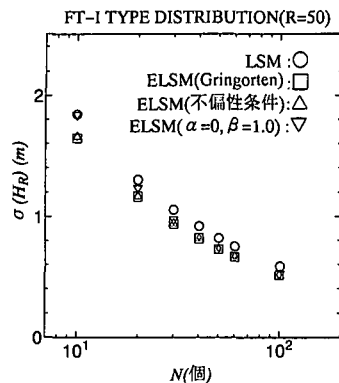
図-4 は、50 年確率波高の標準偏差と標本数 N との関係を示したものである。母集団の分布関数は、FT-I 型分布で母数の値はそれぞれ $A=1.39, B=4.5$ である。図中の○印は、従来の最小二乗法 (LSM) による結果で、プロットング公式は Gringorten の公式を用いている。□印は順序統計量の分散を考慮した母数推定法 (ELSM) で計算した結果で、プロットング公式は Gringorten の公式を用いた結果である。△印は、ELSM によるものでプロットング公式は、式 (29) の不偏性条件式による α および β の値を用いた時のものである。▽印は、同じく ELSM によるものでワイブル公式を用いた結果である。

この図より、従来の最小二乗法 (LSM) による結果より、順序統計量の分散を考慮した母数推定法の方が、約 10% 程度標準偏差が小さく、推定誤差が小さいことを示している。また、Gringorten の公式を用いた結果と不偏性条件式より得られた結果は、ほとんど同一の標準偏差を示している。さらに、バイアスが生じ余り適切でないと考えられているワイブル公式を用いても、順序統計量の分散を考慮すれば、標本数が 20 個以上では従来の最小二乗法によるよりも標準偏差が小さくなることを示している。

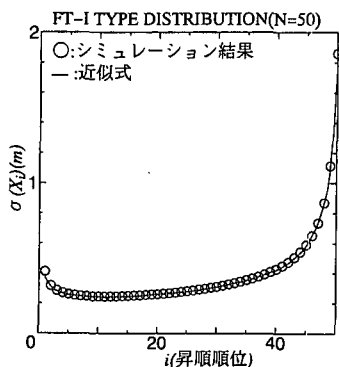
一般に母数推定法の良否は、ある与えられた標本数に



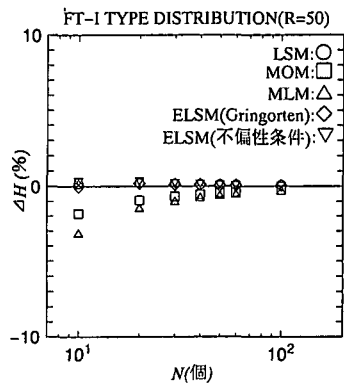
図一 順序統計量 x_i の標準偏差の変化 (FT-I 型分布, $N=20$)



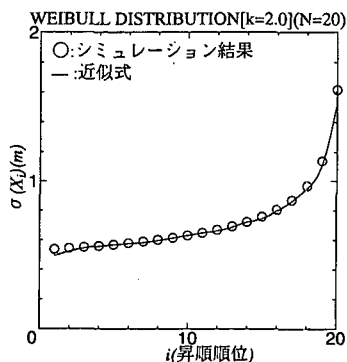
図四 50 年確率統計量の標準偏差と標本数 N との関係



図二 順序統計量 x_i の標準偏差の変化 (FT-I 型分布, $N=50$)



図五 50 年確率統計量のバイアスの比較 (FT-I 型分布)



図三 順序統計量 x_i の標準偏差の変化 (ワイブル分布, $N=20$)

対して、バイアスの絶対値が十分に小さく、確率統計量の推定誤差が小さいという、不偏性および有効性の良否で判断されることが多い。図一5は、各種母数推定法による50年確率統計量のバイアスの変化を示したものである。母数分布関数は、FT-I型分布で $A=1.1$, $B=4.7$ に

対する結果である。○印は従来の最小二乗法 (LSM) による結果、□印は積率法 (MOM) によるもの、△は最尤法 (MLM) によるもの、◇印および▽は順序統計量の分散を考慮した推定法 (ELSM) による結果で、それぞれ Gringorten 公式および不偏性条件式を用いたものである。

最尤法および積率法による結果は、標本数が少ない時には有意なバイアスが存在しているが、Gringorten 公式を用いた最小二乗法や本研究の ELSM 法ではバイアスはほとんど存在せず、標本数 N が10個から20個でも1%以下に留まっている。

一方、50年確率統計量の標準偏差を比較したのが、図一6である。この図によると、最尤法による標準偏差が最も小さく、従来の最小二乗法による結果が最も標準偏差が大きくなっている。順序統計量の分散を考慮した推定法は、最小二乗法や積率法よりも推定誤差は小さくなっているものの、最尤法による結果よりもやや大きくなっている。推定される確率統計量のバイアスと標準偏差の両方の結果を総合すると、本研究で得られた分散を考慮する推定法は、確率統計量の標準偏差については、最尤法にやや劣るが、計算が簡単に行えること、および

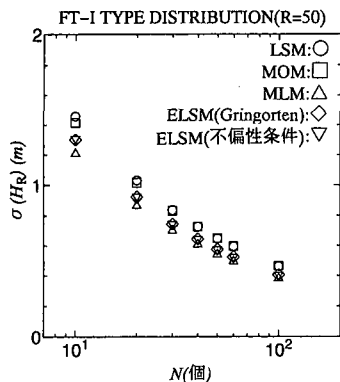


図-6 50年確率統計量の標準偏差の比較 (FT-I型分布)

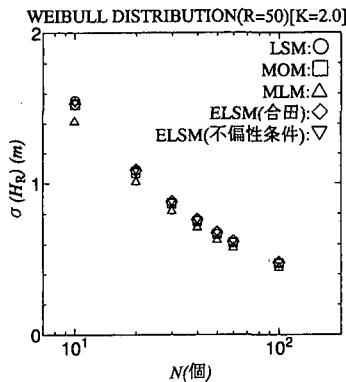


図-8 50年確率統計量の標準偏差の比較 (ワイブル分布)

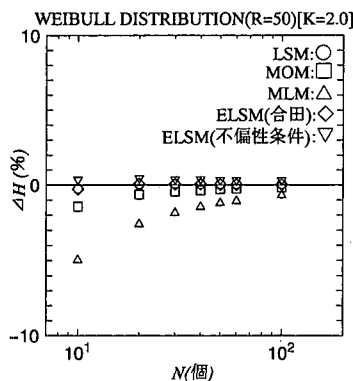


図-7 50年確率統計量のバイアスの比較 (ワイブル分布)

バイアスがほとんどなく、この点については最尤法よりも優れていると言える。

図-7は、ワイブル分布($A=4.36$, $B=0.75$, $k=2.0$)に対する各種母数推定法の50年確率統計量のバイアスの変化を示している。この場合も同様に、最尤法が最もバイアスの絶対値が大きく、最小二乗法や本研究の推定法がバイアスが小さい。50年確率統計量の標準偏差については、図-8に示すように最尤法が最も小さく、その他の方法による結果には有意な差はほとんど見られない。このことは、山口ら(1995)の3母数のワイブル分布に関する数値シミュレーション結果と異なり、その原因については不明である。

6. 結 論

順序統計量の分散を考慮した新たな母数推定法を提案し、数値シミュレーションによりその有効性を検討した結果、以下の事柄が明かとなった。

(1) 漸近理論を用いて順序統計量の分散を簡便に評価する式を得た。この評価式の妥当性を数値シミュレーションにより検討した結果、FT-I型分布およびワイブル分布共によく一致することが確かめられた。本研究では漸近理論を用いているが、標本数が20程度以上でかなりよく一致することが分った。

(2) 順序統計量の分散を考慮した母数推定法を提案し、その精度を検証するために、確率統計量のバイアスおよび標準偏差を他の母数推定法による結果と比較した。本研究の母数推定法は、FT-I型分布の場合には従来の最小二乗法や積率法と比べて10%程度精度が良いことが確かめられた。しかしながら、ワイブル分布については、最尤法を除く他の推定法と同程度の精度であった。

(3) 不偏性の条件式を理論的に示し、プロットング公式の改良を計った。その結果、Gringorten公式や合田による公式と数値は多少異なっているが、標本数によらずバイアスがほとんど0であり、その妥当性が確かめられた。

参 考 文 献

- 合田良実 (1988): 極値統計におけるプロットング公式ならびに推定値の信頼区間に関する数値的検討, 港湾研究所報告, 第27巻, 第1号, pp. 31-92.
- 山口正隆・畑田佳男・中村雄二・大木泰憲 (1995): 波高の極値統計解析における母数推定法の比較, 海岸工学論文集, 第42巻, pp. 231-235.
- Gringorten, I. I. (1963): A plotting rule for extreme probability paper, J. Geophys. Res., Vol. 68, No. 3, pp. 813-814.
- Rao, C. R. (1977): Linear Statistical Inference and Its Applications, John Wiley and Sons, 奥野忠一他訳, 統計的推測とその応用, 東京図書, 568 p.