海岸工学論文集，第 47 巻（2000）
土木学会，216-220

漸近理論を用いた重み付最小二乗法ならびに最適合分布の評価法

泉　宮　尊　司*・岡　本　佳　世**

1. 緒　　言

設計波高および極値分布関数を精度よく推定することは，構造物の破壊確率をより正確に推定することができるため，海岸工学上極めて重要なことである．これまで設計実務において，計算の簡便さから最小二乗法が用いられている(合田，1990)．しかしながら，従来の最小二乗法による評価法では，順序統計量の分散の変化が採り入れられていないために，確率波や最適合極値分布関数の評価において，誤差を伴なうことが危惧される．そこで本研究では，順序付けられた極値データはもはや独立なデータではなく，相関が存在することに着目し，その分散共分散行列を漸近理論により評価し，順序統計量の同時確率分布を求め，その確率を最大とする母数推定法(WLSM)を開発する．さらに，その同時確率密度関数より赤池の情報量規準 AIC を算定し，これを最小とする分布関数を最適合分布関数として採用する手法を提案し，その妥当性について，モンテカルロシミュレーションにより検討する．

2. 漸近理論を用いた重み付最小二乗法

これまでの極値統計解析における母数推定法として，最小二乗法 (LSM) や最尤法 (MLM) などが用いられているが，前者は簡便でバイアスがない推定が可能であるという長所を有している反面，最適合分布関数を評価する最適な方法がないという欠点がある．一方，最尤法は推定精度は最も高いが，標本数が小さい場合には負のバイアスが生じることや非線形方程式を解かなければならないため，計算が複雑であるという欠点がある．本研究では，これら 2 つの推定法の長所を併せ持ち，最尤法の欠点である計算の複雑性を取り除いた母数推定法を開発する．

本研究の母数推定法は，泉宮・斎藤(1998)が示した順序統計量 x_i と x_j の同時確率分布が，漸近的に 2 次元正規分布に従うことを利用する．この関係を N 次元に拡張して考えると，順序統計量 x_i の期待値からの偏差 ϵ_i は，

$$\epsilon_i = x_i - E(x_i) = x_i - Ay_i - B \quad \cdots\cdots (1)$$

多次元正規分布に漸近することから，以下の確率密度関数に漸近することが示される．

$$p(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{\frac{N}{2}} (\det \Sigma)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{M\theta})^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{M\theta})\right\} \quad \cdots (2)$$

ここに，N はデータ数，y_i は順序統計量 x_i に対応する基準化変量，Σ は順序統計量の分散共分散行列であり，

$$\Sigma = E[(\boldsymbol{x}-\boldsymbol{M\theta})(\boldsymbol{x}-\boldsymbol{M\theta})^t] \quad \cdots\cdots (3)$$

である．M は，基準化変量 y_i で表される行列で，

$$M = \begin{pmatrix} y_1, & 1 \\ y_2, & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ y_N, & 1 \end{pmatrix} \quad \cdots\cdots (4)$$

と表される．また，ベクトル $\boldsymbol{x}$ および母数ベクトル $\boldsymbol{\theta}$ は次式で表される．

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_N)^t, \ \boldsymbol{\theta} = (A, B)^t \quad \cdots\cdots (5)$$

分散共分散行列 Σ は，泉宮・斎藤 (1998) が示しているように，母分布関数の真の分散 σ^2 に比例するので，

$$\Sigma = \sigma^2 \Sigma_N \quad \cdots\cdots (6)$$

$$\det \Sigma = \sigma^{2N} \det \Sigma_N \quad \cdots\cdots (7)$$

と表される．ここに，Σ_N は各要素が σ^2 で基準化された分散共分散行列である．上の関係式を式（2）に代入すると，

$$p(\boldsymbol{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \frac{1}{\sqrt{\det \Sigma_N}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{M\theta})^t \boldsymbol{\Sigma}_N^{-1}(\boldsymbol{x}-\boldsymbol{M\theta})\right\} \quad \cdots\cdots (8)$$

となる．ここで，真の分散 σ^2 が一定値であるとすると，基準化された分散共分散行列 Σ_N は，母数には関係しないので，順序統計量の同時確率密度関数の値を最大にするには，指数部の中の項を最大にすればよい．すなわち，

$$S(\boldsymbol{\theta}) = (\boldsymbol{x}-\boldsymbol{M\theta})^t \Sigma_N^{-1}(\boldsymbol{x}-\boldsymbol{M\theta}) \quad \cdots\cdots (9)$$

を最小にすればよいことなる．この $S(\boldsymbol{\theta})$ を最小化する未知パラメタ $\boldsymbol{\theta}$ は，次式で与えられる．

* 正会員　工博　新潟大学教授　工学部建設学科
**　　　　　　　津幡町役場

$$\hat{\boldsymbol{\theta}}=(\boldsymbol{M}^t\Sigma_N{}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}^t\Sigma_N{}^{-1}\boldsymbol{x} \quad \cdots\cdots(10)$$

上式より，推定される母数は，順序統計量の線形結合で表され，泉宮・斎藤（1997）の一般形となっている．式（9）の表現から分かるように，本母数推定法は順序統計量の分散の違いを考慮した重み付最小二乗法（WLSM：Weighted Least Squares Method）であるといえる．

基準化された共分散行列 Σ_N の各要素 c_{ij} は，泉宮・斎藤（1998）により得られた関係式より，FT-I 型分布の場合には，$i<j$ に対して

$$c_{ij}=\frac{6(N+\alpha+\beta-j)}{N\pi^2(j-\alpha)\ln\left(\frac{i-\alpha}{N+\beta}\right)\ln\left(\frac{j-\alpha}{N+\beta}\right)} \quad \cdots\cdots(11)$$

となり，Weibull 分布の場合には，

$$c_{ij}=\frac{(i-\alpha)/[(\Gamma_{2k}-\Gamma_{1k}^2)N(N+\alpha+\beta-i)]}{k^2\left\{\ln\left(\frac{N+\alpha+\beta-i}{N+\beta}\right)\ln\left(\frac{N+\alpha+\beta-j}{N+\beta}\right)\right\}^{1-1/k}} \quad \cdots\cdots(12)$$

となる．ここに，Γ_{nk} は Γ 関数を用いて，次式で表される．

$$\Gamma_{nk}=\Gamma\left(1+\frac{n}{k}\right),\quad (n=1,2,3\cdots) \quad \cdots\cdots(13)$$

なお，$i>j$ の時には，式（11）および式（12）の右辺のみ i と j を入れ換えた式となる．

3. 情報量規準 AIC を用いた最適合分布の評価

従来の最小二乗法による母数推定法では，母数の推定は簡単である反面，最適合分布関数を評価する適切な方法がないという問題がある．これまで，最適合分布関数の評価法として，相関係数を用いる方法や DOL 基準および REC 基準を満足し，かつ MIR 基準により相関係数の相対残差が最小の分布関数を採用する方法（合田，1990），および SLSC を用いる方法（高棹ら，1986）などが提案されている．しかしながら，いずれも相関係数で表されるため，最良の方法であるとは言えない．そこで本研究では，Kullback-Leibler の情報量を用いて評価された赤池の情報量規準 AIC を用いることにする（坂本，1992）．

$$\mathrm{AIC}=-2l(\theta)-2p \quad \cdots\cdots(14)$$

ここに，p は未知母数の数であり，$l(\theta)$ は対数尤度で，式（8）の対数をとることによって求められる．したがって，順序統計量の漸近分布を用いれば，情報量規準 AIC は次式で与えられる．

$$\mathrm{AIC}=\frac{1}{\sigma^2}(\boldsymbol{x}-\boldsymbol{M}\hat{\boldsymbol{\theta}})^t\Sigma_N^{-1}(\boldsymbol{x}-\boldsymbol{M}\hat{\boldsymbol{\theta}})$$
$$+\ln(\det\Sigma_N)+N\ln(2\pi\sigma^2)+2p \quad \cdots\cdots(15)$$

式（15）において，真の母分布関数の分散 σ^2 の値は厳密には算定できないが，次の不偏推定量で置き換えることにより，分布関数に寄らず一定の値で代替できるという利点がある．

$$\sigma^2=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \quad \cdots\cdots(16)$$

ここに，$\bar{x}$ は極値資料の平均値である．

4. 数値シミュレーションによる検証

4.1 重み付最小二乗法の適用性

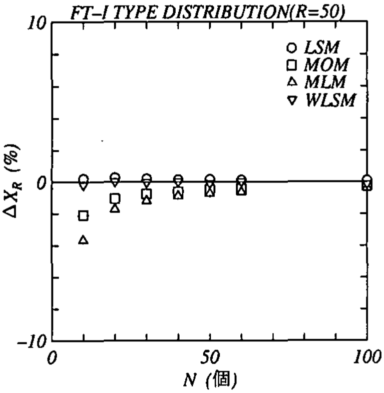
本研究で開発された漸近理論を用いた重み付最小二乗法の適用性を調べるために，モンテカルロ法による数値シミュレーションを行った．対象とした確率分布関数は，FT-I 型分布および Weibull 分布であり，後者については形状母数 k の値を，$k=0.75, 1.0, 1.4, 2.0$ の4種類の分布関数について計算を行っている．標本数 N を，$N=10, 20, 30, 40, 50, 60, 100$ と設定し，5000 回資料を作成して，30 年，50 年，100 年，200 年，500 年および 1000 年再現確率統計量のバイアスおよび分散等を算定した．なお，500 年および 1000 年までの再現確率統計量を算定したのは，特にバイアスに対する性能を調べるためである．

重み付最小二乗法の性能を比較するために，母数推定法として従来の最小二乗法（LSM），積率法（MOM）および最尤法（MLM）についても計算を行い，バイアスおよび標準偏差を推定している．なお，プロッティング公式については，FT-I 型分布に対して，バイアスが殆どないことが確認されている Gringorten 公式（Gringorten, 1963）を，Weibul 分布に対しては合田（1988）による修正 Petruskas and Aagaard 公式を用いている．

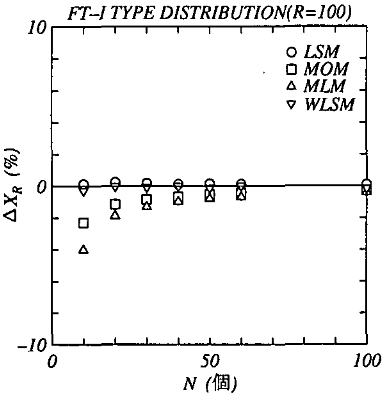
図—1 は，FT-I 型分布（$A=1.39, B=4.5$）に対する 50 年確率統計量のバイアスを各母数推定法ごとに示したものである．この図において，LSM は従来の最小二乗法，MOM は積率法，MLM は最尤法であり，WSLM は重み付最小二乗法である．バイアスについては，重み付最小二乗法および従来の最小二乗法による推定値では，殆ど存在しないが，最尤法および積率法では標本数が 50 以下で，有意な負のバイアスが存在していることが分かる．

図—2 は，同じく FT-I 型分布（$A=1.39$, B$=4.5$）に対する，100 年確率統計量のバイアスを示したものである．**図—1** の結果と同様に，WLSM および LSM には殆どバイアスがないのに対して，MOM および MLM では有意な負のバイアスが見られる．200 年以上の確率統計量のバイアスについても調べた結果，WSLM および LSM では 1000 年確率値に対しても殆どバイアスがないことが確認された．したがって，FT-I 型分布に対してはプロッティング公式として Gringorten 公式の有効性が改めて確認された．

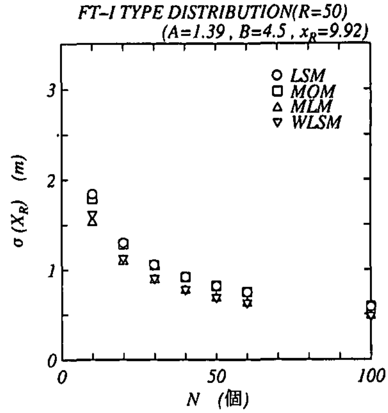
一方，R 年確率統計量の標準偏差を見てみると，**図—3**



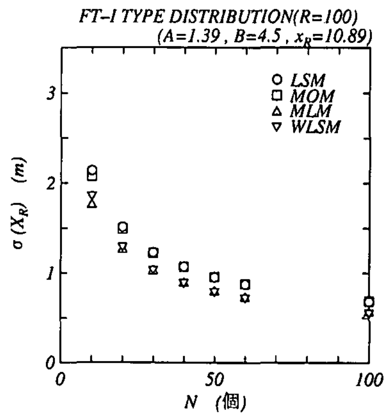
図—1 50年確率統計量のバイアスの比較（FT-I 型分布）



図—2 100年確率統計量のバイアスの比較（FT-I 型分布）



図—3 50年確率統計量の標準偏差（FT-I 型分布）



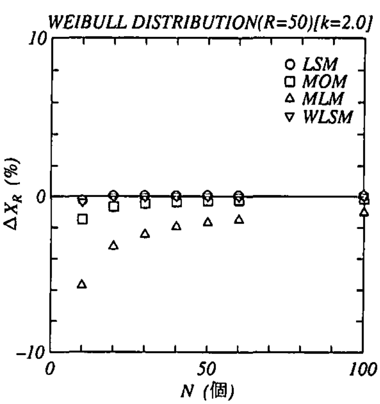
図—4 100年確率統計量の標準偏差（FT-I 型分布）

(R=50年) および図—4 (R=100年) に示すように，本研究で開発されたWLSMの標準偏差は，データ数が30以上であれば最尤法（MLM）のそれにほぼ等しくなり，推定精度が高いことが分かる．それに対して，従来の最小二乗法（LSM）および積率法（MOM）では，R年確率統計量の標準偏差が約1割程度大きくなっている．

図—5および図—6は，Weibull分布 (k=2.0) の場合の50年確率統計量のバイアスの変化を示している．この分布関数では，合田（1988）による修正Petruskas and Aagaard公式を用いている．これらの図から分かるように，LSMおよびWLSMともにバイアスが殆どなく，図には示していないが，200年，500年および1000年確率統計量に対してもバイアスが殆どないことが確認された．

Weibull分布の場合のR年確率統計量の標準偏差については，図—7および図—8に示すようにFT-I型分布ほど差異はないが，WLSMによる推定結果は，MLMについで標準偏差が小さく，N=30以上ではMLMの結果とほぼ同一の値となっている．

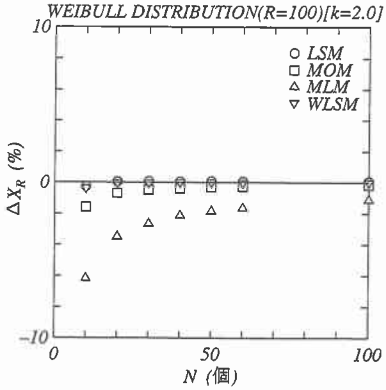
以上のように，本研究で開発されたWLSMは，バイアスの面ではMOMやMLMより優れており，R年確率統計量の標準偏差では従来のLSMより優れていることが示された．



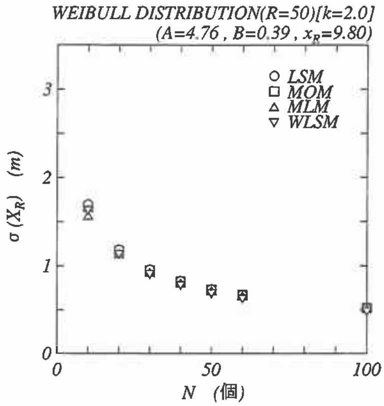
図—5 50年確率統計量のバイアスの比較（Weibull分布）

4.2 最適合分布関数の評価性能

最適合分布関数の評価法として，本研究で開発された



図—6 100年確率統計量のバイアスの比較（Weibull分布）

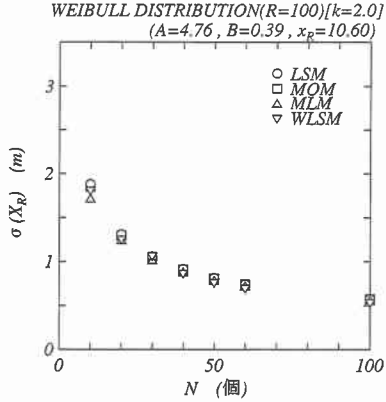


図—7 50年確率統計量の標準偏差（Weibull分布）

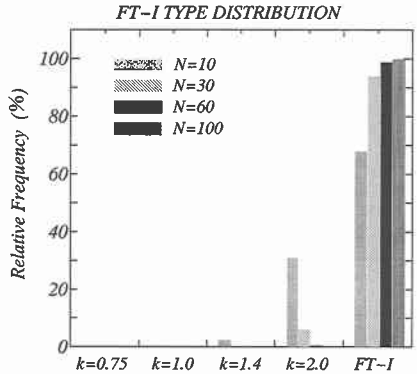
順序統計量の同時確率分布の漸近分布より導かれるAIC規準の評価性能を調べるために，モンテカルロシミュレーションを行った．FT-I型分布およびWeibull分布（k=1.0，2.0）の三種類の分布関数に従うデータをそれぞれ5000ケース作成し，FT-I型分布およびWeibull分布（k=0.75，1.0，1.4，2.0）の5種類の分布関数に当てはめて，WLSMにより母数を計算し，AICを評価した．

真の分布関数がFT-I型分布の場合，**図—9**に示すようにデータ数がX=30で，90%以上真の母分布関数を探し当てる性能を有し，N=100ではほぼ100%の正解率となっている．このように，FT-I型分布に対して評価性能が高いのは，前節で示したようにR年確率統計量の標準偏差が従来のLSMに比べて1割ほど小さく，その分だけ評価性能が高くなった結果であると考えられる．

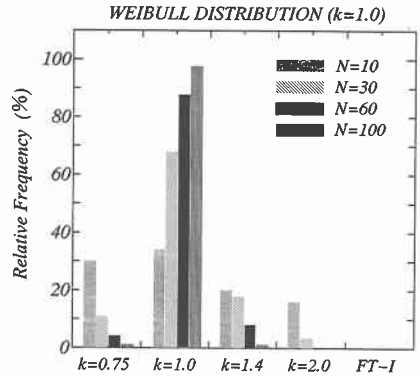
図—10および**図—11**には，真の分布関数がWeibull分布で，形状母数がそれぞれk=1.0およびk=2.0のときの適合度評価結果を示している．**図—10**のk=1.0のときには，他の形状母数のWeibull分布に適合する確率が



図—8 100年確率統計量の標準偏差（Weibull分布）



図—9 AICによる適合度評価（FT-I型分布）

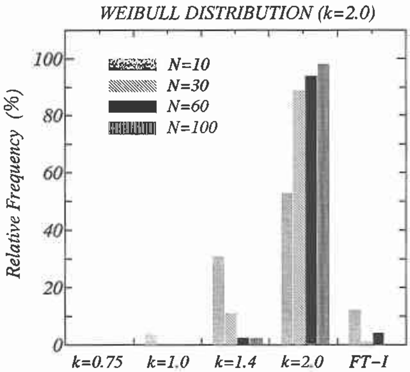


図—10 AICによる適合度評価（Weibull分布，k=1.0）

やや大きいが，それでもデータ数N=60でおおよそ90%の正解率となっている．また，形状母数k=2.0の場合には，データ数N=30で90%近い正解率となっている．

5. 結　　論

本研究では，順序統計量の同時確率密度関数の漸近分



図—11　AICによる適合度評価（Weibull分布，$k=2.0$）

布を用いて，その確率を最大とする母数推定法WLSMおよび情報量規準AICによる適合度判定法を開発し，数値シミュレーションによりその有効性を検討したところ，以下の事柄が明らかとなった。

（1）　R年確率統計量の標準偏差は，最尤法（MLM）に次いで小さく，標本数が30以上あれば最尤法と同程度の精度が得られることが明かとなった。

（2）　バイアスについては，FT-I型分布およびWeibull分布に対して，WLSMは従来の最小二乗法(LSM)と同じく殆ど0であり，最尤法や積率法よりも優れている。

（3）　WLSMによる母数推定法では，情報量規準AICの評価も行列演算で簡便に評価でき，標本数が60以上であれば約90%以上の高い精度を持って，真の確率分布関数を見出すことができることが分かった。

参考文献

泉宮尊司・斎藤雅弘（1997）：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，海岸工学論文集，第44巻，pp. 181-185.

泉宮尊司・斎藤雅弘（1998）：極値統計解析における不偏性条件ならびに漸近理論による信頼区間の推定，海岸工学論文集，第45巻，pp. 206-210.

合田良実（1988）：極値統計解析におけるプロッティ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，港湾技術研究所報告，第27巻，第1号，pp. 31-92.

合田良実(1990)：港湾構造物の耐波設計—波浪工学への序説—，鹿島出版会，333 p.

坂本慶行（1992):カテゴリカルデータのモデル分析，共立出版，pp.1-15.

高棹琢馬・宝　馨・清水　章（1986）：琵琶湖流域水文データの基礎的分析，京大防災研年報，第29号，B-2，pp. 157-171.

Gringorten, I. I. (1963): A plotting rule for extreme probability paper, J. Geophys. Res. Vol. 68, No. 3, pp. 813-814.