波高の極値統計解析における最尤法の適用性の検討

山口正隆*・前川隆海**・大木泰憲***・中村雄二****

1. 緒　　言

波高の極値統計解析で用いられる各種理論確率分布に対する母数推定法のうち，最尤法は不偏性，有効性，一致性，十分性という性質を近似的に満足するとともに，censoringをうけた部分年最大値資料にも適用できる理論的に優れた方法であるが，計算量が大きいため，その特性を包括的に検討した研究事例は少ない．

そこで，本研究では，母集団分布を3母数Weibull分布，GEV分布，3母数対数正規分布およびGumbel分布に指定した場合の大規模なモンテカルロシミュレーション結果の最尤法を用いた解析に基づいて，確率波高のbiasおよび標準偏差推定値に及ぼす形状母数と標本資料数の影響を詳細に調査することにより，最尤法，統計量のbias補正・分散推定法であるjackknife法，漸近的分散推定法である情報行列法の適用性を明らかにする．また，Gumbel分布の場合には，censoring(資料採択率)の影響を検討する．

2. 確率分布と母数・分散推定法

(1) 確率分布

3母数Weibull分布，GEV分布，3母数対数正規分布(ひずみ係数正)およびGumbel分布の非超過確率$F(x)$はそれぞれ次式で表される．

$$\left.\begin{array}{l} F(x)=1-\exp[-\{(x-B)/A\}^k]\,;\,B<x<\infty \\ F(x)=\exp[-\exp\{1-(x-B)/kA\}^k] \\ \qquad\qquad x<B+kA\,;\,k>0,\,x>B+kA\,;\,k<0 \\ F(x)=(1/\sqrt{\pi})\int_{-\infty}^{y}\exp(-y^2)dy \\ \qquad\qquad y=k\cdot\log\{(x-B)/A\}\,;\,B<x<\infty \\ F(x)=\exp[-\exp\{-(x-B)/A\}]\,;\,-\infty<x<\infty \end{array}\right\} \quad\cdots\cdots(1)$$

ここに，x：確率変数，A：尺度母数，B：位置母数，k：形状母数，である．

* 正会員　工博　愛媛大学教授　工学部環境建設工学科

** 学生会員　愛媛大学大学院　理工学研究科

***　工修　東亜建設工業株式会社

****　工修　株式会社　ハザマ

(2) 母数推定法

最尤法は標本資料に対する確率密度関数$f_i(x_i)$の積で定義される尤度関数Lを最大化するように母数を決定する方法であるが，下限のcensoringを考慮した年最大値資料に対しては，尤度関数の定義を，

$$L=\{N!/(N-r)!\}\{F(x_c)\}^{(n-r)}\prod_{i=1}^{r}f_i(x_i) \quad\cdots\cdots(2)$$

と修正する必要がある(たとえば，Phien・Emma, 1989)．ここに，N：資料総数，r：censoringされていない資料数，$N-r$：censoringされた資料数，x_c：censoringが行われた下限基準値，である．上式は非超過確率が$F(x_c)$である$N-r$個のcensoringされた資料が尤度関数に及ぼす影響を考慮したものである．波浪追算に基づいて年最大波高資料を作成する場合，波浪追算対象気象擾乱を勢力の強いものに限定することが多い．そのため，低波高をもつ年最大波高資料が的確に評価されない場合があるので，censoringの影響を考慮した極値解析も必要になる．

なお，下限censoringの影響を考慮する原始的な方法の1つは，基準値以下の資料を切り捨てた年最大波高資料を改めて資料相当期間の年最大波高資料とみなす方法であり，この方法の精度も検討する．

(3) jackknife法の適用

統計量のbias補正法や分散(標準偏差)推定法として，近年，jackknife法(Miller, 1974)などの計算機依存型resampling手法が提案されている．jackknife法は，N個の標本資料から得られる統計量$\hat{\theta}$と1個の標本資料を順次除いたN組の標本(標本資料数$N-1$個)を用いて得られる統計量$\theta_{*i}(i=1\sim N)$の演算によって，統計量のbiasの補正や分散の推定を行う方法である．jackknife法を多数回適用する場合には，膨大な計算時間を要するので，ここでは，その適用を$N\leqq200$の場合に限定する．

(4) 漸近的分散推定法

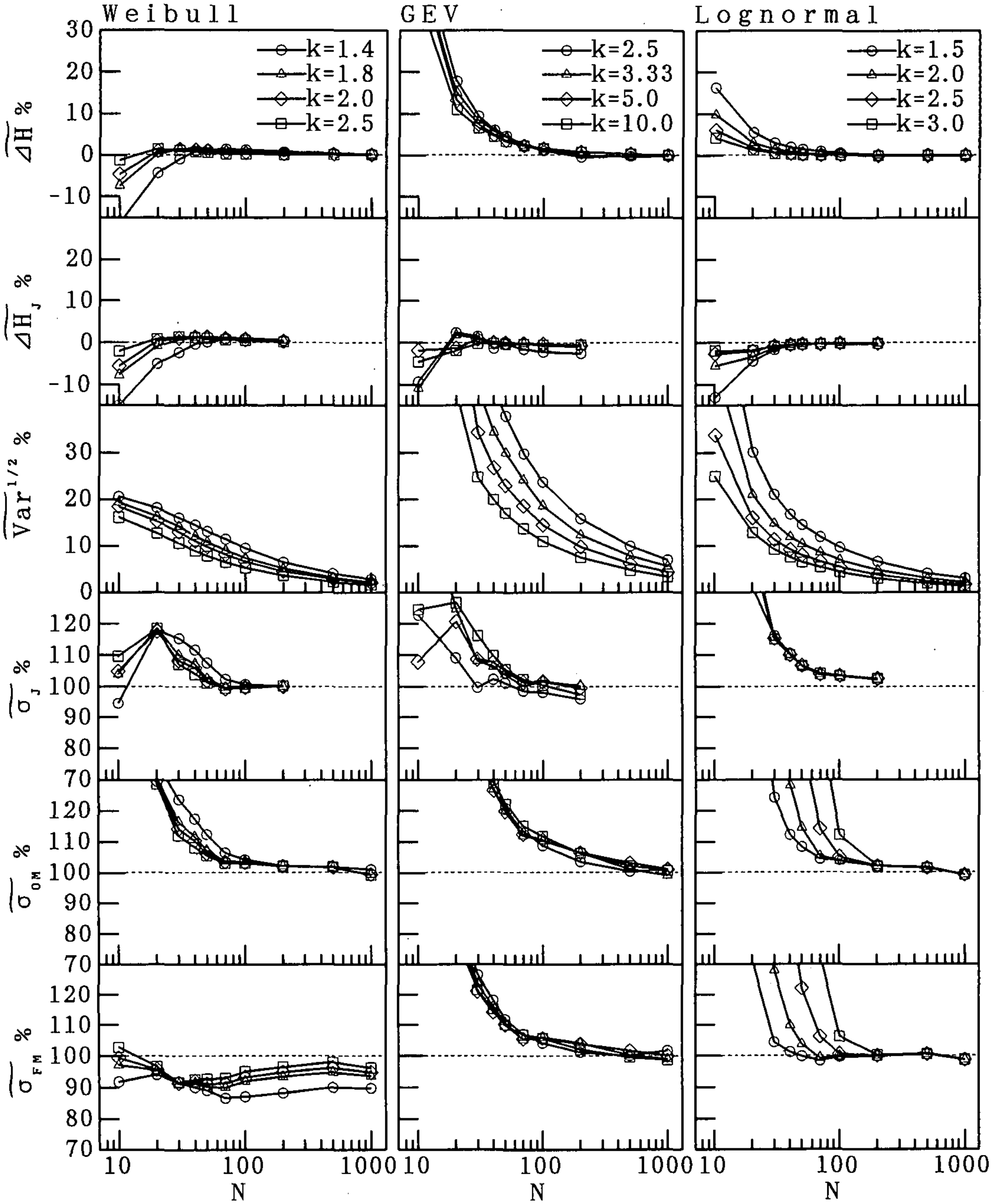
最尤法による場合，対数尤度関数の母数に関する2階微分値を要素とする行列の逆行列が母数の分散・共分散行列に相当する性質を利用して，統計量の分散を推定できる．2階微分値そのものを要素とする行列は観測情報行列，2階微分値の期待値を要素とする行列はFisher情

報行列と呼ばれる．本研究で取り扱う理論確率分布に対する観測・Fisher 情報行列は上・下限の censoring を受けた場合を含めて導かれている（たとえば，Phien・Emma, 1989）．しかし，資料採択率が $\nu<1$ の場合，Fisher 情報行列に基づく表示式の数値計算は面倒であるので，本研究では検討対象としない．

3. シミュレーションの方法

シミュレーションの手順はつぎのようである．

① 確率分布の inverse form 式中の非超過確率に一様乱数を順次与えて，任意個数 N の標本を抽出し，最尤法による母数の推定，確率波高とその bias ΔH の算出，jackknife 法による bias 補正後の bias ΔH_J と分散 σ_J^2 の



図—1 3母数分布に対する誤差統計量（1）

計算ならびに観測・Fisher 情報行列法による分散 σ^2_{OM}, σ^2_{FM} の演算を行う．また，censoring を考慮する場合には，生成したN個の標本のうち，上位の νN 個を用いて同様の計算を行い，σ^2_{FM} を除く諸量を求める．

② 同じ手順を $M(=5000)$ 回繰り返し，①の諸量の平均値および確率波高の標準偏差 $Var^{1/2}$ を求める．

③ $N=10\sim1000$ の間の 10 通りの場合に，①，②の手順を繰り返す．

シミュレーションは，3 母数分布の場合には形状母数を 4 種類(位置母数，尺度母数は一定，$\nu=1.0$)，Gumbel 分布の場合には資料採択率を 4 種類変えて行う．すなわち，入力母数は，Weibull 分布では $A=4.0$ m, $B=1.0$ m, $k=1.4$, 1.8, 2.0, 2.5, GEV 分布では $A=1.0$ m, $B=5.0$ m, $k=2.5$, 3.33, 5.0, 10.0, 対数正規分布では $A=3.5$ m, $B=2.5$ m, $k=1.5$, 2.0, 2.5, 3.0, Gumbel 分布では $A=1.39$ m, $B=4.5$ m, $\nu=1.0$, 0.9, 0.8, 0.7 である．また，確率波高は再現期間 $R=50$, 100, 200, 500, 1000 年に対して推定する．

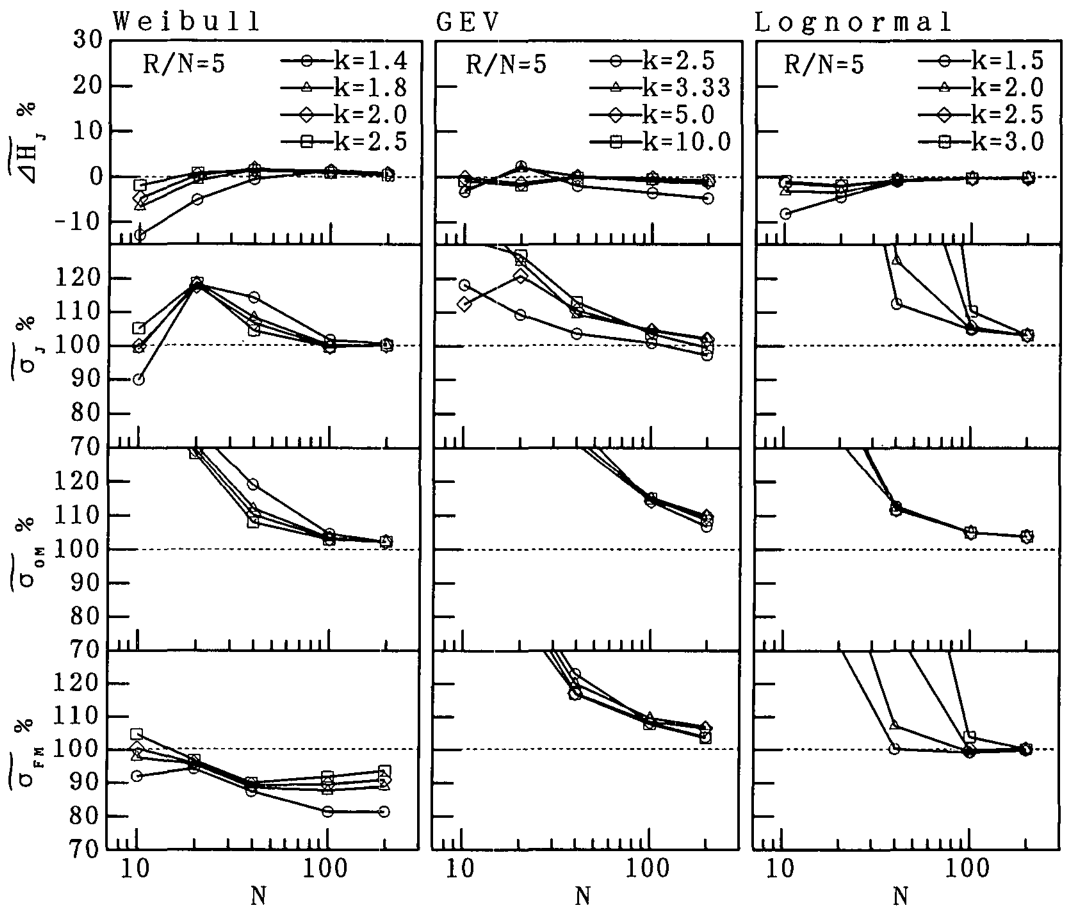
4. シミュレーション結果の考察

図—1 は，3 母数分布である Weibull 分布，GEV 分布および対数正規分布の場合に，100 年確率波高の真値で無次元化した bias $\widetilde{\Delta H}$, $\widetilde{\Delta H_J}$ と $\widetilde{Var}^{1/2}$，および $Var^{1/2}$ で無次元化した標準偏差推定値，$\tilde{\sigma}_J$, $\tilde{\sigma}_{OM}$, $\tilde{\sigma}_{FM}$ に対する％表示値（誤差統計量）と資料数Nの関係を形状母数 k をパラメータとして示したものである．3 分布に共通して，つぎの特徴が見出される．

① 形状母数が小さいほど，分布の裾が広がるので，$Var^{1/2}$ が増加する．その結果，確率波高および $Var^{1/2}$ の推定精度が低下するが，資料数の増大とともに，それらの推定精度が向上する．

② 最尤法と jackknife 法を組み合わせた方法の適用範囲は，bias の観点では $N\geq20\sim30$, $Var^{1/2}$ の推定精度の観点では $N\geq40\sim50$ とするのが望ましい．

③ 観測情報行列法は資料数の増加とともに過大評価から適正評価へ移行する．その適用限界は，確率分布の種類や形状母数の値（対数正規分布の場合）に依存する



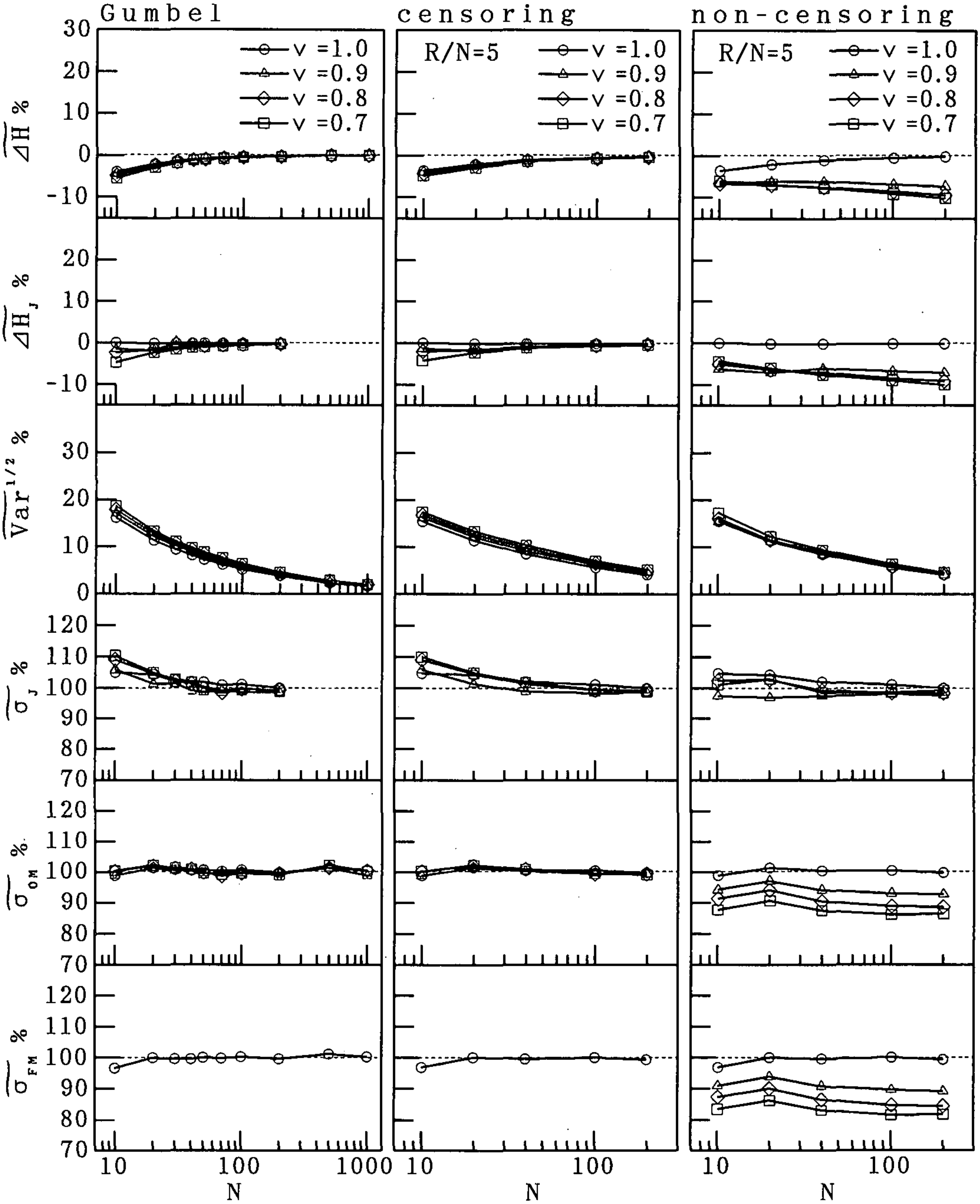
図—2 3 母数分布に対する誤差統計量（2）

が，おおむね $N\geqq 50$ である．Fisher 情報行列法も，資料数によらず標準偏差を10%程度過小評価する Weibull 分布の場合を除いて，小標本に対して標準偏差を過大評価する．ただし，観測情報行列法より小さい標準偏差推定値を与えるので，観測情報行列法より推定精度が高い．その適用範囲は $N\geqq 40\sim 50$ と考えられる．

ついで，各分布に対する個別の検討結果はつぎのようである．すなわち，Weibull 分布の場合，

① 最尤法は，対数尤度関数の母数に関する1階微分式から明らかなように，$k\leqq 1$ では解を与えない．精度的には，最尤法の適用を $k\geqq 1.4$ とするのが望ましい．

② jackknife 法は bias 補正に対して必ずしも有効に機能しない．また，分散推定法の精度は，分布幅の狭い $k=1.4$ の場合を除いて，形状母数にあまり依存しない．



図—3 Gumbel 分布に対する誤差統計量

GEV分布の場合,

① 最尤法は$k\leqq 3$では適切な解を与えない.これは積率解の範囲が$k>3$であることから傍証される.

② jackknife法はbias補正に対して有効に機能する.また,確率波高や標準偏差の推定精度に及ぼす形状母数の影響は,jackknife法に基づく標準偏差推定法($N<50$の場合)を除いて,小さい.

対数正規分布の場合,

① jackknife法は$N\geqq 20$でbias補正に有効である.

② jackknife法による標準偏差の推定精度は$N\geqq 30$では形状母数にほとんど依存しない.一方,情報行列法は形状母数が大きいほど標準偏差を過大評価する.

図—2は再現期間R年と資料数Nとの比が$R/N=5$の場合に対する誤差統計量と資料数の関係を形状母数をパラメータとして示したものであり,資料は$(R,N)=(50, 10), (100, 20), (200, 40), (500, 100), (1000, 200)$の5ケースに対応する.また,誤差統計量はそれぞれの再現期間に対する値で無次元化されている.これによると,いずれの分布においてもjackknife法によるbias補正後のbiasは,$N>20\sim 40$であれば,かなり小さいことや,jackknife法による標準偏差の推定誤差は$N\geqq 40$では10~15%以下であることから,最尤法とjackknife法を組み合わせた方法は$N\geqq 40$であれば,有効であると考えられる.この結果は$R/N=10$の場合も同様である.観測情報行列法およびFisher情報行列法(Weibull分布を除く)の適用範囲は,$R/N=10$の結果を含めて考えると,Weibull分布では$N\geqq 40$, GEV分布および対数正規分布では$N\geqq 50$とみなされる.ただし,前述のように,Fisher情報行列法の精度がより高い.また,Weibull分布の場合,Fisher情報行列法は標準偏差を過小評価する傾向にあるが,その誤差は15%以内である.

図—3はGumbel分布の場合の誤差統計量と資料数の関係を資料採択率をパラメータとして示したものである.左図はcensoringの影響を考慮した場合の再現期間100年に対する結果,中央の図はcensoringの影響を考慮した場合で$R/N=5$(資料の組合せは図—2と同じ)に対する結果,右図はcensoringの影響を考慮しない場合の$R/N=5$に対する結果を表す.まず,左図および中央の図や$R/N=10$に対す結果(図省略)より,つぎのことが見出される.

① 最尤法は,資料採択率によらず,資料数が小さいほど負のbiasを生じる.biasは,$\nu=1$の場合には,jackknife法の適用によってほぼ解消されるが,$\nu<1$の場合には,biasはあまり減少しない.

② 資料採択率が減少するほど,$Var^{1/2}$は増加する.また,jackknife法の標準偏差推定精度は良好であるが,$N=10$で$\nu\leqq 0.8$の場合には,やや低下する.

③ 観測情報行列法の精度は資料採択率や資料数によらず非常に高い.これはFisher情報行列法の場合($\nu=1$の場合のみ)も同様である.

また,右図よりつぎのことが云えよう.

① censoringの影響を考慮しない場合,最尤法は確率波高を10%程度過小評価する.この傾向はjackknife法の適用によっても補正されない.

② $Var^{1/2}$は資料採択率の減少とともに若干増加する.また,jackknife法は,資料数や資料採択率によらず,標準偏差を±5%の誤差内で推定する.

③ 情報行列法は資料数によらず,資料採択率が低いときほど,標準偏差を過小評価する.この傾向はFisher情報行列法の場合により著しい.

5. 結 語

最尤法とjackknife法を併用する確率波高および標準偏差推定法ならびに情報行列法に基づく標準偏差推定法は2母数分布であるGumbel分布の場合,資料数や資料採択率によらず良好な精度を与えるが,3母数分布の場合,その適用範囲を$N\geqq 50$とすることが推奨される.

参 考 文 献

Miller, R. G. (1974): The jackknife-a review, Biometrica, Vol. 61, No. 1, pp. 1-15.

Phien, H. N. and F. T. S. Emma (1989): Maximum likelihood estimation of the parameters and quantiles of the general extreme-value distribution from censored samples, Jour. Hydrol., No. 105, pp. 139-155.