

極値統計解析における不偏性条件式ならびに 漸近理論による信頼区間の推定

泉 宮 尊 司*・斎 藤 雅 弘**

1. 緒 言

ある限られた標本数の極値データから確率波高を精度よく推定するためには、バイアスがなく、しかも推定される確率統計量の分散がより小さい推定法を用いることが望ましい。これまで数多くの数値シミュレーションが行われ、これらの検証や信頼区間の経験式が提案されている(合田, 1988, 1990, 山口ら, 1995, 1996)。しかしながら、これらの数値シミュレーションによる検討も、母分布関数や母数の値が異なると、結果も相違することが指摘されており(山口ら, 1996)、一般的かつ普遍的な結論が出せない状況にある。

一方、プロットング公式に関しても、有限の標本数に対してバイアスのない公式が確立されたとは言い難く、母関数が異なればバイアスが生じることも指摘されている。さらに、信頼区間の推定については、厳密にかつ理論的に評価しようとする、最小二乗法では順序統計量の同時確率分布を用いて、やや複雑な数値計算をしなければならないので、実用に供せられるような公式を得ることはかなり困難である。

そこで本研究では、まずプロットング公式を用いた場合の不偏性条件式を示すと共に、波浪極値解析によく用いられている FT-I 型分布および Weibull 分布に対して、バイアスのない公式を提案する。また、積率法および最小二乗法を用いた場合の R 年確率統計量の分散を、それらの分布に対して漸近理論を用いて理論的に求め、それらの精度についてモンテカルロ法により検討する。

2. 最小二乗法における不偏性条件式

最小二乗法における不偏性条件式については、泉宮・斎藤 (1997) により簡単に述べられているが、ここではその計算法について少し詳しく述べることにする。最小二乗法を用いた場合には、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。

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

ここに、 g_i および h_i は次式で表される。

$$g_i = k_i, \quad h_i = \frac{1}{N} - \bar{y} k_i \quad \cdots \cdots (2)$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad k_i = \frac{y_i - \bar{y}}{\sum_{j=1}^N (y_j - \bar{y})^2} \quad \cdots \cdots (3)$$

ここに、 y_i は標準化変量であり、FT-I 型分布の場合には、

$$y_i = -\ln[-\ln(F_i)], \quad F_i = \frac{i-\alpha}{N+\beta} \quad \cdots \cdots (4)$$

また、Weibull 分布の場合は、

$$y_i = [-\ln(1-F_i)]^{-1/k}, \quad F_i = \frac{i-\alpha}{N+\beta} \quad \cdots \cdots (5)$$

と表される。上式より、 y_i がプロットング公式の α および β の関数となるので、式 (1) における係数 g_i および h_i もそれらの関数となる。式 (1) 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、

$$E[\hat{A}] = \sum_{i=1}^N g_i E[x_i] = \sum_{i=1}^N g_i \mu_i = A \quad \cdots \cdots (6)$$

$$E[\hat{B}] = \sum_{i=1}^N h_i E[x_i] = \sum_{i=1}^N h_i \mu_i = B \quad \cdots \cdots (7)$$

となる。ここに、 $\mu_i = E[x_i]$ は、FT-I 型分布の場合、次式で与えられる。

$$\begin{aligned} \mu_i &= \frac{N!}{(i-1)!(N-i)!} \\ &\times \int_0^1 [A \ln(-\ln t) + B] t^{i-1} [1-t]^{N-i} dt \\ &= A I_i + B \quad \cdots \cdots (8) \end{aligned}$$

ここに、

$$\begin{aligned} I_i &= \frac{N!}{(i-1)!(N-i)!} \\ &\times \int_0^1 [-\ln(-\ln t)] t^{i-1} [1-t]^{N-i} dt \quad \cdots \cdots (9) \end{aligned}$$

である。Weibull 分布の場合には、形状母数を k とすると、 I_i は

$$\begin{aligned} I_i &= \frac{N!}{(i-1)!(N-i)!} \\ &\times \int_0^1 [-\ln(1-t)]^{1/k} t^{i-1} [1-t]^{N-i} dt \quad \cdots \cdots (10) \end{aligned}$$

となる。式 (8) を式 (6) および式 (7) に代入すると、

$$A = \sum_{i=1}^N g_i \mu_i = A \sum_{i=1}^N g_i I_i + B \sum_{i=1}^N g_i \dots (11)$$

$$B = \sum_{i=1}^N h_i \mu_i = A \sum_{i=1}^N h_i I_i + B \sum_{i=1}^N h_i \dots (12)$$

なる関係式が得られる。定義式 (2) より、 $\sum_{i=1}^N g_i = 0$ 、 $\sum_{i=1}^N h_i = 1$ であるので、上の関係式が成立するためには、

$$\sum_{i=1}^N g_i I_i = 1, \quad \sum_{i=1}^N h_i I_i = 0 \dots (13)$$

でなければならない。これらの関係式を満たす α および β は、次の誤差ノルム L を用いて、

$$L^2 = \left(\sum_{i=1}^N g_i I_i - 1 \right)^2 + \left(\sum_{i=1}^N h_i I_i \right)^2 \dots (14)$$

を最小とするものと等価であるので、式 (14) を用いてパイアスのないプロットング公式を得ている。

3. 漸近理論による信頼区間の推定

3.1 積率法による確率統計量の分散の推定

積率法による確率統計量の分散の推定は、FT-I 型分布に対して、Gumbel (1957) によって行われている。いま、積率法によって推定された R 年確率統計量を $\hat{x}_R$ とすると、その分散 $\sigma^2(\hat{x}_R)$ は、

$$\sigma^2(\hat{x}_R) = \frac{\sigma^2}{N} \{ 1 + 0.8885(y_R - \gamma) + 0.6687(y_R - \gamma)^2 \} \dots (15)$$

である。ここに、 y_R は x_R に関する基準化変数であり、 σ^2 は母集団の分散、および $\gamma = 0.5772$ はオイラー数である。式 (15) の関係は、漸近的に正規分布となる平均値 $\bar{x}$ および標準偏差 s に関する量の分散が

$$\sigma^2(a\bar{x} + bs) = \frac{\sigma^2}{N} \{ a^2 + ab\sqrt{\beta_1} + b^2(\beta_2 - 1)/4 \} \dots (16)$$

で評価されることから求められている (Gumbel, 1957)。ここに、

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \dots (17)$$

であり、 $\sqrt{\beta_1}$ および β_2 は、それぞれひずみ度 (skewness) および尖鋭度 (kurtosis)、 a および b はある定数である。式 (16) より、確率統計量の分散を推定するには、1 次から 4 次までのモーメントを求めればよい。

Weibull 分布の場合には、 n を正の整数とすると、

$$E[(x-B)^n] = A^n = A^n \Gamma\left(1 + \frac{n}{k}\right) \dots (18)$$

なる関係が成立することから、容易に 4 次モーメントまでを求めることができる。それらの結果を用いて、

$$\sqrt{\beta_1} = \frac{(\Gamma_{13k} - 3\Gamma_{11k}\Gamma_{12k} + 2\Gamma_{11k}^3)}{(\Gamma_{12k} - \Gamma_{11k}^2)^{3/2}} \dots (19)$$

$$\beta_2 = \frac{(\Gamma_{14k} - 4\Gamma_{13k}\Gamma_{11k} + 6\Gamma_{12k}\Gamma_{11k}^2 - 3\Gamma_{11k}^4)}{(\Gamma_{12k} - \Gamma_{11k}^2)^2} \dots (20)$$

を得る。ここに、 Γ_{1nk} は、 Γ 関数を用いて、

$$\Gamma_{1nk} = \Gamma\left(1 + \frac{n}{k}\right), \quad (n=1, 2, 3, \dots) \dots (21)$$

である。積率法による確率統計量 $\hat{x}_R$ は、

$$\begin{aligned} \hat{x}_R &= \hat{A}y_R + \hat{B} \\ &= \bar{x} + \frac{(y_R - \Gamma_{11k})}{(\Gamma_{12k} - \Gamma_{11k}^2)^{1/2}} s \dots (22) \end{aligned}$$

と表されることから、式 (16) より $\hat{x}_R$ の分散は、

$$\begin{aligned} \sigma^2(\hat{x}_R) &= \frac{\sigma^2}{N} \left\{ 1 + \frac{(y_R - \Gamma_{11k})}{\sqrt{\Gamma_{12k} - \Gamma_{11k}^2}} \sqrt{\beta_1} \right. \\ &\quad \left. + \frac{(y_R - \Gamma_{11k})^2}{(\Gamma_{12k} - \Gamma_{11k}^2)} \frac{(\beta_2 - 1)}{4} \right\} \dots (23) \end{aligned}$$

となる。ここに、 $\sqrt{\beta_1}$ および β_2 は、式 (19) および式 (20) を用いる。

3.2 最小二乗法による確率統計量の分散の推定

最小二乗法による確率統計量の分散は、推定母数が式 (1) で表されること、および不偏性条件を満たすことを用いて、評価することができる。いま、 R 年確率統計量の真値および推定値を、それぞれ x_R および $\hat{x}_R$ とすると、

$$x_R = Ay_R + B, \quad \hat{x}_R = \hat{A}y_R + \hat{B} \dots (24)$$

と表されることから、式 (1) を代入し、その両辺の期待値をとると、

$$x_R = \sum_{i=1}^N (g_i y_R + h_i) E(x_i), \quad \hat{x}_R = \sum_{i=1}^N (g_i y_R + h_i) x_i \dots (25)$$

となる。この関係を用いると、推定値 $\hat{x}_R$ の分散は、

$$\begin{aligned} \sigma^2(\hat{x}_R) &= E[(\hat{x}_R - x_R)^2] \\ &= E \left[\sum_{i=1}^N (g_i y_R + h_i) (x_i - E(x_i)) \right]^2 \\ &= \sum_{j=1}^N \sum_{i=1}^N (g_i y_R + h_i) (g_j y_R + h_j) \\ &\quad \times E[(x_i - E(x_i))(x_j - E(x_j))] \\ &= \sum_{j=1}^N \sum_{i=1}^N (g_i y_R + h_i) (g_j y_R + h_j) C_{ij} \end{aligned} \dots (26)$$

となる。ここに、

$$C_{ij} = E[(x_i - E(x_i))(x_j - E(x_j))] \dots (27)$$

であり、 x_i および x_j に関する共分散行列である。これを厳密に評価するには、順序統計量 x_i および x_j の確率分布を用いて数値積分を行えばよいが、漸近理論を用いると、比較的簡単に評価することができる。

いま、順序統計量 x_i および x_j ($i < j$) の同時確率分布を $p(x_i, x_j)$ とすると、

$$p(x_i, x_j) = \frac{N!}{(i-1)!(j-i-1)!(N-j)!} \times [F(x_i)]^{i-1} [F(x_j) - F(x_i)]^{j-i-1} \times [1 - F(x_j)]^{N-j} f(x_i) f(x_j) \dots (28)$$

と表せる (Rao, 1977)。ここで、順序の入れ替わることのない 1 対 1 対応の付く次の変換を考えて、 ξ_i と ξ_j の漸近分布を考えることにする。

$$\xi_i = F(x_i), \quad \xi_j = F(x_j) \dots (i < j) \dots (29)$$

この変換により、 ξ_i と ξ_j の同時確率密度関数 $\phi(\xi_i, \xi_j)$ は、

$$\phi(\xi_i, \xi_j) = p(x_i, x_j) |J| = \frac{N!}{(i-1)!(j-i-1)!(N-j)!} [\xi_i]^{i-1} \times [\xi_i, \xi_j]^{j-i-1} [\xi_j]^{N-j} \dots (30)$$

となる。ここに、 $|J|$ は Jacobian である。 ξ_i の期待値は、 $i/(N+1)$ となるが、 $N \rightarrow \infty$ に対して、 $p_i = (i-\alpha)/(N+\beta)$ と等しくなるので、次の変換を考える。

$$z_i = \sqrt{N} \left\{ \xi_i - \frac{i-\alpha}{N+\beta} \right\}, \quad z_j = \sqrt{N} \left\{ \xi_j - \frac{j-\alpha}{N+\beta} \right\} \dots (31)$$

この時、 z_i および z_j の同時確率分布は、比例係数を除いた部分は次式となる。

$$\phi(z_i, z_j) \sim \left\{ \frac{z_i}{\sqrt{N}} + p_i \right\}^{i-1} \left\{ \frac{z_j - z_i}{\sqrt{N}} + (p_j - p_i) \right\}^{j-i-1} \times \left\{ 1 - \frac{z_j}{\sqrt{N}} - p_i \right\}^{N-j} \sim \left\{ 1 + \frac{z_i}{\sqrt{N} p_i} \right\}^{i-1} \left\{ 1 + \frac{z_j - z_i}{\sqrt{N} (p_j - p_i)} \right\}^{j-i-1} \times \left\{ 1 - \frac{z_j}{\sqrt{N} (1 - p_i)} \right\}^{N-j} \dots (32)$$

ここに、 $p_j = (j-\alpha)/(N+\beta)$ である。ここで、上式の数式をとり、

$$\log(1-x) = -x - x^2/2 - \dots, \quad (x \ll 1) \dots (33)$$

なる関係を用いると、 $N \rightarrow \infty$ に対して、

$$\ln \phi(z_i, z_j) \sim -\frac{1}{2} \left\{ \frac{z_i^2}{p_i} + \frac{(z_j - z_i)^2}{(p_j - p_i)} + \frac{z_j^2}{(1 - p_i)} \right\} \dots (34)$$

となる。したがって、 z_i と z_j は漸近的に 2 次元正規分布に従うことになる。ここで、 z_i と z_j との共分散行列を (R_{ij}) とすると、式 (34) より、

$$\left. \begin{aligned} \frac{R_{ij}}{|R|} &= \frac{1}{(p_j - p_i)}, & \frac{R_{ji}}{|R|} &= \frac{1}{p_i} + \frac{1}{(p_j - p_i)} \\ \frac{R_{ii}}{|R|} &= \frac{1}{(1 - p_i)} + \frac{1}{(p_j - p_i)} \end{aligned} \right\} \dots (35)$$

なる関係が成立するため、 $|R| = p_i(1-p_i)(p_j-p_i)$ を得る。また、 R_{ij} および R_{ii} は次式で与えられる。

$$R_{ij} = p_i(p_j - p_i), \quad R_{ii} = p_i(1 - p_i) \dots (36)$$

この (z_i, z_j) に関する共分散行列より、(x_i, x_j) に関する共

分散行列 C_{ij} は、変換の Jacobian を用いて次式となる。

$$C_{ij} = \frac{p_i(1-p_i)}{Nf(x_{ip})f(x_{jp})}, \quad (i < j) \dots (37)$$

$i > j$ の時は、上式の右辺のみを i と j 入れ換えた式となる。FT-I 型分布の場合には、 $i < j$ に対して、

$$C_{ij} = \frac{6\sigma^2(N+\alpha+\beta-j)}{N\pi^2(j-\alpha)\ln\left[\frac{j-\alpha}{N+\beta}\right]\ln\left[\frac{j-\alpha}{N+\beta}\right]} \dots (38)$$

となる。また、Weibull 分布の場合には、同じく $i < j$ に対して、

$$C_{ij} = \frac{A^2(i-\alpha)}{\left(N(N+\alpha+\beta-i)k^2 \left\{ \ln\left[\frac{N+\alpha+\beta-i}{N+\beta} \right] \right\} \times \ln\left[\frac{N+\alpha+\beta-j}{N+\beta} \right] \right)^{1-1/k}} \dots (39)$$

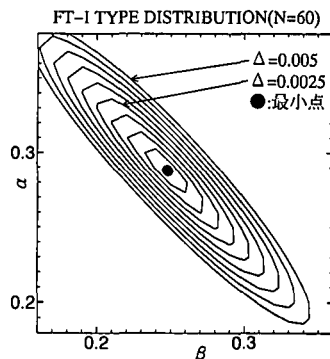
となる。ここに、 $A^2 = \sigma^2 / (\Gamma_{12k} - \Gamma_{11k}^2)$ である。

以上より、式 (26) とそれぞれの分布に対して、式 (38) あるいは式 (39) の関係を用いることによって、 $\hat{x}_k$ の分散を評価することができる。なお、部分極値資料の場合には、全極大値資料の数を N_T 、採択した極値資料の数を N とするとき、式 (26) の (i, j) を (i, j) = $N_T - N + 1$ から N_T まで足し算を行えばよく、この評価式は部分極値資料に対しても適用可能である。

4. 数値シミュレーションによる検証

4.2 バイアスなしのプロットング公式

バイアスなしのプロットング公式および確率統計量の分散の理論式の適用性を調べるために、モンテカルロ法により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、標本数 N を $N=10, 20, 30, 40, 50, 60, 100$ と設定し、5000 個の資料を作成して、確率統計量のバイアスおよび分散を算定した。



図一 誤差ノルム Δ の分布 ($N=60$)

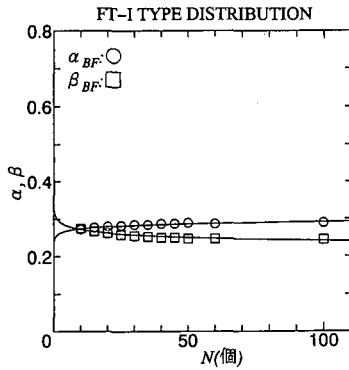


図-2 不偏性条件を満たす α および β の値 (FT-I 型分布)

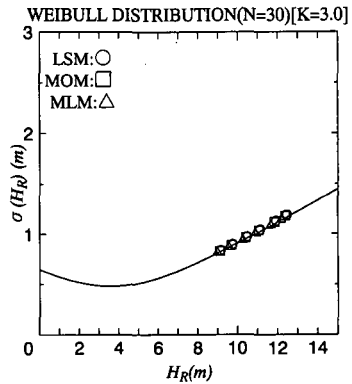


図-5 R 年確率統計量の標準偏差の比較 (Weibull 分布, $k=3.0$)

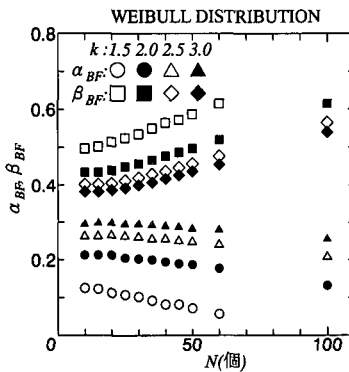


図-3 不偏性条件を満たす α および β の値 (Weibull 分布)

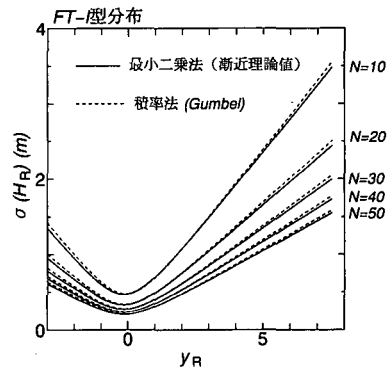


図-6 R 年確率統計量の標準偏差の漸近理論値との比較 (FT-I 型分布)

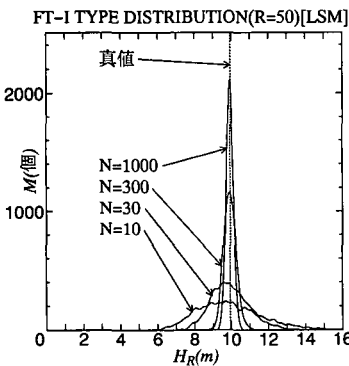


図-4 50 年確率統計量の標本数 N に対する分布

図-1 は、式 (14) で定義される誤差ノルムの等値線を示したものである。この図より、誤差ノルムはほぼ楕円分布しており、その極小値を格子探索法により求めて、バイアスなしのプロットング公式を得ている。

このようにして得られた α および β の値を、標本数 N に対してプロットしたものが、図-2 および図-3 である。図-2 は、FT-I 型分布の結果であり、図-3 は Weibull 分布に対するものである。FT-I 型分布の場合に

は、標本数に殆ど寄らずほぼ一定の値となっていることが分かる。これに対して、Weibull 分布では、形状母数 k の値が小さいほど、標本数に依存して有意に変化している。また、形状母数 k の値が小さいほど、 α の値は小さく、かつ β の値は大きくなっていることが分かる。

4.2 確率統計量の分散の推定結果

一般に推定された母数を用いて、例えば 50 年確率統計量を推定した場合、それは確定的に定まるものではなく、漸近的に正規分布する確率変数となる。数値シミュレーションによって、それを確認したものが図-4 である。この図は、FT-I 型分布の標本を用いて、最小二乗法により 50 年確率統計量を推定した結果である。標本数 $N=30$ 程度でやや非対称性が認められるものの正規分布に近くなっていることが分かる。したがって、その推定精度はほぼその分散あるいは標準偏差で計ることが可能である。

図-5 は、母分布関数が Weibull 分布で、標本数が $N=30$ の時の R 年確率統計量の標準偏差を示したものである。図中の実線は、式 (23) の積率法の標準偏差の漸近

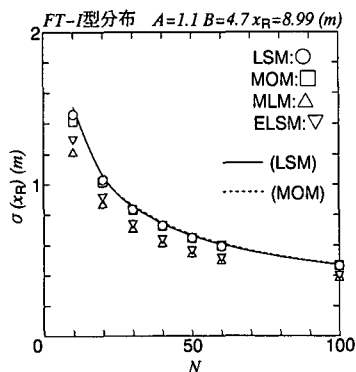


図-7 50年確率統計量の標準偏差の漸近理論値の比較 (FT-I型分布)

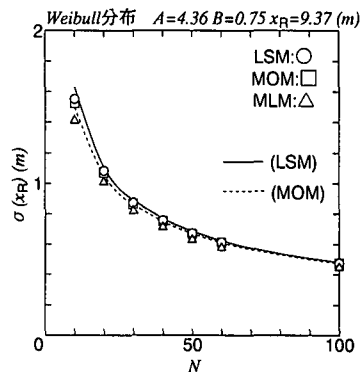


図-8 50年確率統計量の標準偏差の漸近理論値の比較 (Weibull分布)

理論値である。この理論値と数値計算結果との一致度は高く、負のバイアスを伴う最尤法 (MLM) の結果も、それぞれ H_R の値が小さい方へ移動しても、その曲線上にあるのは非常に興味深いものである。

図-6 は、FT-I 型分布 ($A=1.39$, $B=4.5$) の場合の R 年確率統計量の標準偏差の漸近理論値を示したものである。実線が最小二乗法の理論値、破線は積率法のそれである。この条件では、両者に大きな差異は見られないが、基準化変量 y_R の値が3以上で、最小二乗法の標準偏差の方が多少小さくなっている。

FT-I 型分布 ($A=1.1$, $B=4.7$) に対する 50 年確率波高 ($H_{50}=8.99$ m) の標準偏差のシミュレーション結果と理論値とを比較したものが、図-7 である。最小二乗法 (LSM), 積率法 (MOM), 最尤法 (MLM) および泉宮・斎藤 (1997) による推定法 (ELSM) の結果も同時に示している。この分布の場合には、最小二乗法および積率法の理論値に殆ど差異はなく、数値計算とほぼ同様の結果となっている。この図には示していないが、プロットング公式に Gringorten 公式を用いると、式 (26) による標準偏差は約 1% 小さくなり、Weibull 公式を用いた場合には約 2~3% 小さく、よりシミュレーション結果と一致することが確認された。

図-8 は、Weibull 分布 ($A=4.36$, $B=0.75$, $H_{50}=9.37$ m) に対する 50 年確率波高の標準偏差を比較したものである。この場合には、積率法の方が標準偏差はやや小さい結果となっており、その理論値との一致度も良好である。一方、最小二乗法の結果は、標本数が小さい場合に多少の差異が認められるが、 $N=30$ 以上ではよく一致していると言える。なお、プロットング公式に Weibull 公式を用いても、殆ど同じ標準偏差の値が得られており、これについては FT-I 型分布とは異なる結果であった。

5. 結 論

極値統計解析における不偏性条件を示し、有限個の標

本に対してもバイアスのない公式を提案するとともに、高次モーメントおよび漸近理論を用いて、積率法および最小二乗法による確率統計量の標準偏差を評価する理論式を得ることによって、以下の事柄が明かとなった。

(1) 不偏性条件を満たすバイアスのないプロットング公式は、FT-I 型分布の場合には、標本数にあまり寄らずほぼ一定値となるが、Weibull 分布の場合には、形状母数 k の値が小さいほど有意に変化する。

(2) Weibull 分布に対する積率法の信頼区間の評価式は、標本数が 30 以上であればシミュレーション結果と殆ど一致することから、簡便さも含めて適用性が高いと言える。

(3) 最小二乗法を用いた場合の確率統計量の分散は、漸近理論によりかなり精度よく評価することができ、同じく標本数が 30 以上であれば、実用に供せられるものである。また、この評価法は部分極値資料に対しても、僅かな修正で適用することができる利点を有している。

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.
- 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.
- 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.
- 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.
- 山口正隆・大木泰憲・前川隆海 (1996): 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, 海岸工学論文集, 第 43 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.