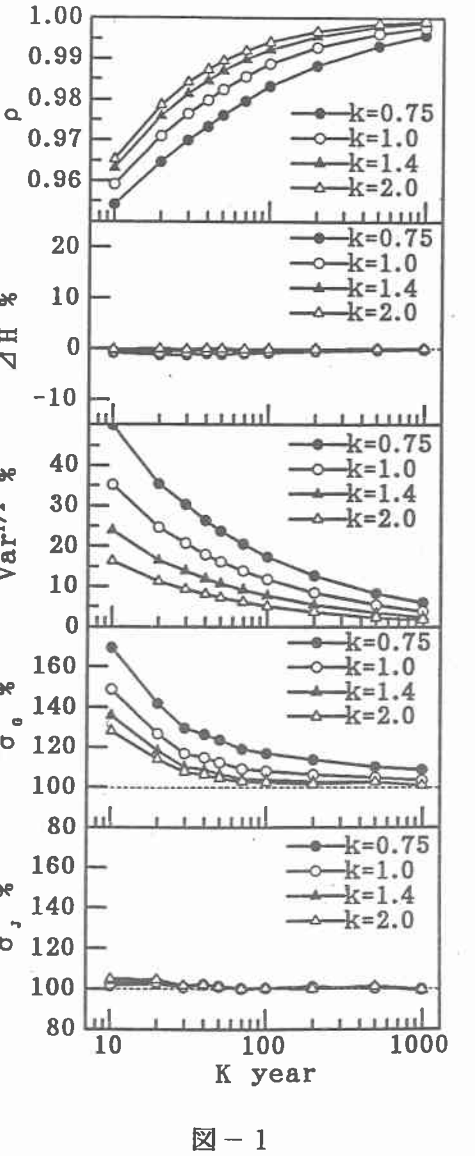
最小２乗法に基づく極値統計解析システムの拡張と精度の検討

愛媛大学工学部　正員　山口正隆　　　愛媛大学大学院　学生員○前川隆海

１．はじめに：合田により開発された「最小２乗法に基づく極値統計解析システム」（現行システム）は候補分布であるWeibull分布の形状母数の上下限をk=2.0および0.75としているために、本来k＞2.0およびk＜0.75をとるべき資料の適合度および再現期待値の推定精度に問題を有する。同様に、FT-Ⅱ型分布では形状母数の上下限をk=10.0および2.5としており、k=10.0とk=∞（Gumbel分布に相当）の間の分布形状の変化が大きい。そこで、本研究では、現行システムの適用範囲を拡張した修正システムを提案し、入力条件を広範囲に設定した場合のモンテカルロシミュレーション結果に基づいて現行システムおよび修正システムの精度を、確率分布の形状母数、データ採択率および極大値資料の年平均発生数の影響という観点から検討する。

２．研究内容：(1)現行システム；このシステムは、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、k=2.5、3.33、5.0、10.0とする４種類のFT-Ⅱ型分布の合計９種類の確率分布を資料のあてはめに使用し、相関係数基準やMIR基準に従って最適確率分布を選択するとともに、確率波高およびその信頼区間の推定結果を得るものである。(2)修正システム；このシステムは候補分布の数をGumbel分布およびWeibull分布ではk=0.5～100、FT-Ⅱ型布ではk＞2.5の範囲で任意の数とした場合に、合田のプロッティングポジション公式を用いた非超過確率表示に基づく確率波高の推定法、相関係数基準およびjackknife法に基づく確率波高の分散推定法を組み合わせる方法である。(3)シミュレーションの方法；①母集団の確率分布と母数およびデータ採択率$\nu=N/N_T$（N：採択資料数、N_T：資料総数）、年平均発生数$\lambda=N_T/K$(K：年数）を指定し、R年確率波高（R=50、100、200、500、1000年）の真値を求める。②確率分布のinvese formに一様乱数を与えてN_T個の標本を発生させ、そのうち上位N個を抽出する。③候補分布の母数を母分布の母数に等しくした場合（既知母分布）あるいは上記の範囲内で任意に設定した場合（未知母分布）について最小２乗法により母数を推定し、相関係数ρ（適合度指標）、R年確率波高H_Rと合田式に基づく確率波高の分散σ_G^2およびjackknife法に基づく分散σ_J^2を求める。未知母分布の場合には、相関係数最大の候補分布を最適分布とし、その分布についての諸量を求める。④対象年数K=10、20、30、40、50、70、100の７ケースあるいは200、500、1000年を加えた10ケースの場合に、②および③の手順を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biasΔH、標準偏差$\mathrm{Var}^{1/2}$を算定する。また、合田式による分散σ_G^2およびjackknife法による分散σ_J^2の平均値などの各種誤差統計量を算定する。(4)結果の考察；図－１は母分布を形状母数k=1.4の３母数Weibull分布とした場合の年最大値資料（K=N=N_T）に対する誤差統計量（既知母分布を仮定）と資料年数Kの関係を、４種類の形状母数の場合に示す。誤差統計量は上から、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$\Delta\tilde{H}$、標準偏差$\widetilde{\mathrm{Var}}^{1/2}$、および100年確率波高の$\mathrm{Var}^{1/2}$で無次元化した合田式およびjackknife法による100年確率波高の平均分散の平方根$\sqrt{\sigma_G^2}$、$\sqrt{\sigma_J^2}$



図－１

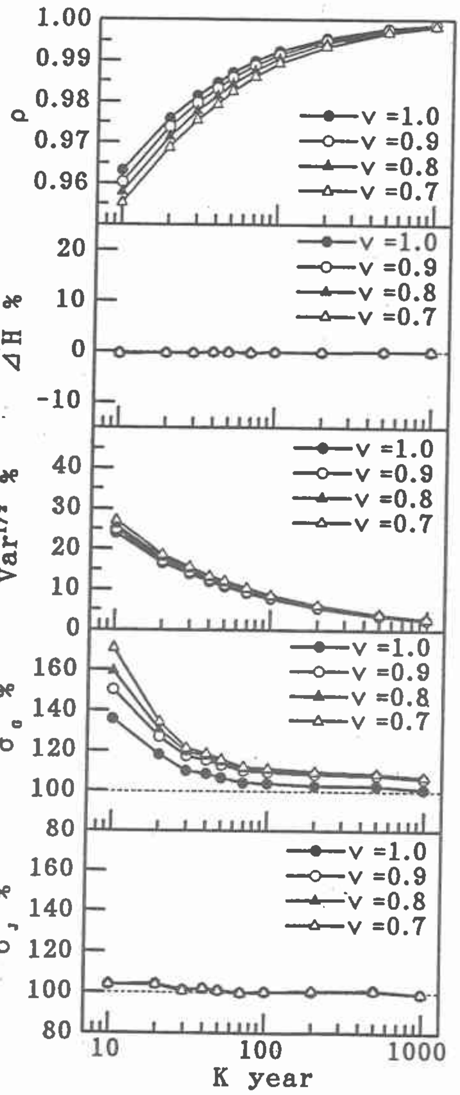
の％表示値を表す。この図によれば、形状母数が小さいほど、分布の裾が広がるので、相関係数 ρ は小さくなり、標準偏差 $\widetilde{\mathrm{Var}}^{1/2}$ は大きくなるが、bias $\Delta\tilde{H}$ は、kにかかわらずほとんど生じない。合田式は形状母数が小さいほど標準偏差を過大評価するが、資料年数の増加とともに、その精度は向上する。jackknife法に基づく標準偏差推定値は、資料年数が小さい場合に若干大きい値をとるものの、それ以外では $\widetilde{\mathrm{Var}}^{1/2}$ とほぼ一致する。図－2は誤差統計量に及ぼすデータ採択率の影響をk=1.4のWeibull分布に基づく年最大値資料（既知母分布）の場合に示したものである。データ採択率が高いほど ρ は大きく、$\widetilde{\mathrm{Var}}^{1/2}$ は小さくなり、合田式による標準偏差推定の精度は良くなる。一方、bias $\Delta\tilde{H}$ とjackknife法に基づく標準偏差推定値の誤差は非常に小さく、データ採択率や資料年数の影響は見られない。図－3は年平均発生数 λ の影響を検討するため、母分布、候補分布ともにk=1.4の3母数Weibull分布とした場合の極大値資料（データ採択率 ν=0.25）に対する誤差統計量を示したものである。年平均発生数の増加とともに、ρ は大きくなり、$\widetilde{\mathrm{Var}}^{1/2}$ は小さくなる。年平均発生数の増加は資料数の増加を表すから、この結果は合理的である。また、bias $\Delta\tilde{H}$ とjackknife法に基づく標準偏差推定値には、年平均発生数や資料年数の影響はほとんど見られず、本システムの精度はきわめて良好であることがわかる。一方、合田式による標準偏差推定値は標準偏差を過大評価するが、年平均発生数が大きい場合にはこの傾向はかなり改善される。

3．まとめ：本研究の検討結果はつぎのように要約される。(1)既知母分布ケースに対する本システムは、母分布が正側に長く裾をひく形状をとる場合（Weibull分布ではk＜0.75、FT-Ⅱ型分布ではk＜3.33）を除いて、形状母数、データ採択率および年平均発生数によらず、確率波高および標準偏差を高い精度で評価する。

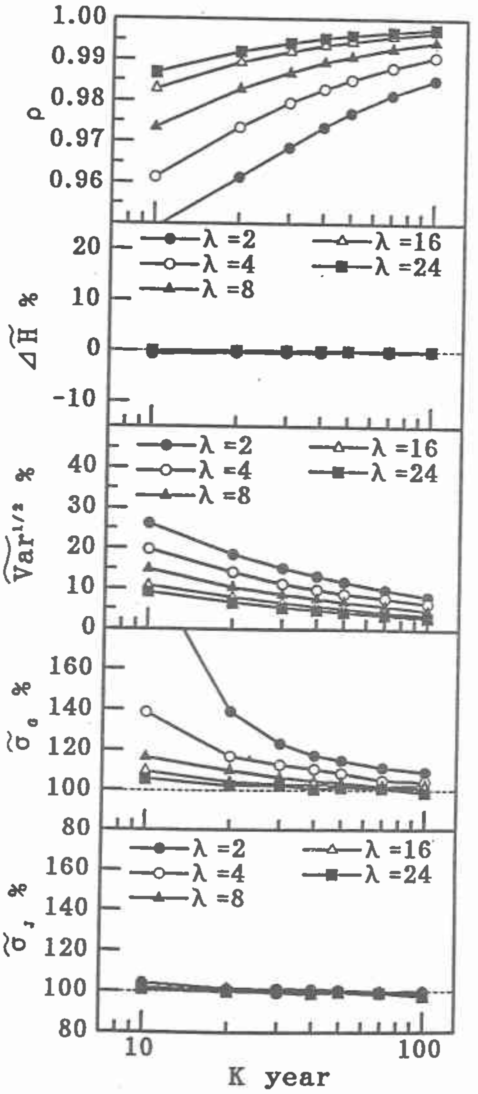
(2)データ採択率、年平均発生数が増大するほど、適合度は向上し、$\mathrm{Var}^{1/2}$ は減少することから、確率波高の推定精度の向上には資料数の増加が必須である。

(3)jackknife法の標準偏差推定に対する誤差は数％にとどまっていることから、jackknife法に基づく分散推定法は実用上十分な適用性をもつといえる。

(4)合田式は標準偏差を過大評価する傾向にある。とくに、年最大値資料の場合に、この傾向は著しいが、資料年数の増加に伴って推定精度は向上する。



図－2



図－3