

極値分布関数の適合度評価に関する検討

藤部 文昭*

要旨

極値統計解析において、複数の確率分布の中からデータに適合するものを選ぶという方法が使われることがある。このようなアプローチの有効性を調べるため、Gumbel分布などに従う母集団からデータを抽出し、これに5つの極値分布関数のそれぞれを当てはめて推定した再現期待値のバイアスと不確定幅(推定幅)を、モンテカルロ・シミュレーションで評価した。その結果、パラメーター2個の分布(Gumbel分布など)は、母集団の確率分布と異なると再現期待値にバイアスが生ずるが、パラメーター3個の分布はほとんどバイアスが無く、それよりもデータの確率変動による推定幅のほうがはるかに大きかった。このことから、パラメーター3個の分布同士に関しては適合度を評価するメリットが小さいと考えられる。次に、適合度の基準として使われるSLSCや、推定幅の評価に使われるジャックナイフ法の特徴を、モンテカルロ・シミュレーションで調べた。その結果、SLSCはデータ年数に依存し、従って、SLSCによる適合性判定のしきい値として定数を使うべきではないことが結論された。また、ジャックナイフ法は再現期待値の推定幅をほぼ正しく評価できるが、推定幅の大きさと適合度とは別であり、ジャックナイフ法を適合度の判定に使うのは適切でないことが指摘された。

1. はじめに

極値統計解析で再現期間や再現期待値を計算する際、当てはめる分布関数を、何らかの基準に基づいてデータに最も良く適合するように決める方法が取られることがある。もともと極値統計では、年極値にGumbel分布、あるいはこれの拡張である一般化極値分布(GEV=generalized extreme value distribution)を当てはめることが主流だった。その理論的背景は、データがすべて同じ確率分布に従い、互いに独立で、多数個あるという条件(以下「同一分布・独立・多数の条件」)を満たせば、その極値がGEVに従うという性質である(Coles 2001; 本城・篠田 2008)。年間の日々の観測値(例えば日降水量)を同一分布・独立・多数の条件を満たす確率変数と見なせ

れば、その年極値はGEVに従うと期待できる。Coles (2001)は、GEVを基本にして様々な発展的手法を解説している。

しかし、現実のデータは同一分布・独立・多数の条件を満たすとは限らず、そのため、年極値がGEVに従う保証はない。このことから、GEV以外にいくつかの確率分布関数を用意し、その中からデータに一番良く適合するものを選ぶという発想が生まれる。適合度の尺度としては、高棹ほか(1986)が提案したSLSC(standard least-squares criterion; 第4節参照)が比較的簡便なものとしてよく使われる。

その一方、適合度に基づいて極値分布関数を選択することの有効性は、必ずしもよく吟味されていない。これには2つの問題がある。1つは、適合度の評価方法としてSLSCが有効かどうかである。葛葉(2010)は「SLSCは非常によくできた客観的規準」であることを認めた上で、SLSCの値がデータ数(統計年数)に依存することを指摘し、SLSCによる適合度判定のしきい値として定数を使うことを批判している。

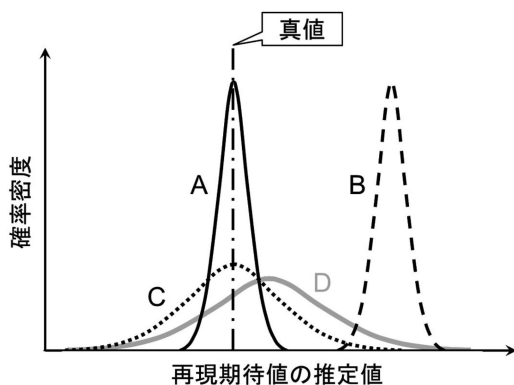
* 気象研究所予報研究部。

—2011年1月11日受領—

—2011年6月15日受理—

もう1つの、より根本的な問題は、そもそも関数の選択をする必要があるのか、という点である。極値統計の目的は、極値発現の背後にあるべき母集団の性質(関数形やパラメーター値)を推定し、これに基づいて極値の再現期間や再現期待値を算出することであり、従って、母集団の確率分布と違う関数を当てはめれば再現期待値の推定値に何がしかのバイアスが生ずるだろう。しかし統計的推定においては、バイアスだけでなく、データ数が有限であることから来る不確実性も重要である。極値統計理論は、年極値の現れ方が確率過程に支配されることを前提にしており、年極値データには年々のランダムな変動(確率変動)が含まれる。そのため、データ数が有限である限り、たとえ正しい(母集団と同じ)関数を当てはめても母集団のパラメーター値が正確に求まるとは限らず、再現期待値の推定値にも相応の不確定幅(以下「推定幅」)が生ずる。そして、もし推定幅がバイアスよりも大きければ、関数の選択をするメリットは乏しいことになるであろう。

第1図は前段落の議論を図にしたものである。母集団が定常ならば、再現期待値も定数であり、それが極値統計で求めたい真値である。しかし、適用する極値分布関数が母集団に一致する保証がないことと、データに確率変動が含まれることにより、データから算出された推定値は真値に一致するとは限らず、何らかの確率分布をする。第1図には、いろいろな極値分布関数を使って推定を行った場合を想定し、それぞれの推定値の確率密度分布を模式的に描いてある。Aはバイアスがなく推定幅も小さい場合、Bはバイアスはあ



第1図 再現期待値の推定値の確率密度分布に関する模式図。A~Dはそれぞれ別の極値分布関数を使って計算した結果を表す。

るが推定幅が小さい場合、Cはバイアスはないが推定幅が大きい場合である。また、Dは若干のバイアスがあるが、それよりも推定幅がはるかに大きい場合である。AとBの比較から分かるように、バイアスが推定幅よりも大きいとき、言い換えると2つの推定値の確率分布に重なりがないときには、極値分布関数の選択が問題になり、適合性の評価が重要な意味を持つ。しかし、CとDを比べた場合のように、推定幅が大きく2つの推定値の確率分布が大幅に重なっていれば、どちらの極値分布関数を使うかは重要ではなく、適合度を評価するメリットは乏しいことになろう(図を見るとCの方がバイアスがないため良いように思えるが、実際の解析場面では真値は分からないことに注意してほしい)。

推定幅を見積もる方法はいくつかあり、ジャックナイフ法はその1つである(第5節参照)。宝・高棹(1988)はSLSCにジャックナイフ法を加味して確率分布関数を選ぶことを提案した。この方法は、中小河川計画検討会(1999)の「中小河川計画の手引き(案)」(以下「手引き」)や気象庁の「異常気象リスクマップ」(以下「リスクマップ」; www.data.kishou.go.jp/climate/riskmap/index.html; 2011年4月25日閲覧)に採用されている(ただし第6節で論ずるように、宝・高棹(1988)の提案と「手引き」「リスクマップ」の方法との間には重要な違いがある)。

本稿では、極値統計解析における極値分布関数の適合度評価に関して、(1)SLSCの特性と、(2)適合度評価の必要性について議論する。まず(2)を取り上げ、モンテカルロ・シミュレーションに基づいて、極値分布関数の違いによる再現期待値のバイアスと、データの確率変動に伴う推定幅を比較する。次に、同じくモンテカルロ・シミュレーションを使ってSLSCの特性を調べ、併せてジャックナイフ法についても議論する。

なお二宮(2011)の質問や、それに対する回答(藤部2011)で論じられたように、極値統計に関しては、そもそもデータは確率変数だと言えるのか、仮にそう言っても母集団は定常なのか・・・といった、より根本的な問題もある。しかし、これらは本稿の範囲を超える事柄である。極値統計手法の多くは、年極値が確率変数であり、定常な母集団に属するという前提に基づいており、以下この前提のもとで議論を進める。

2. 解析方法

2.1 確率分布関数

「手引き」や「リスクマップ」では、Gumbel分布、GEV、平方根指数型最大値分布 (square-root-exponential type maximum distribution, SQRT-ET) を主とし、対数ピアソンIII型分布 (LP3) と対数正規分布 (GNO) を副として各地点に適合する確率分布関数を決めている。本稿ではこれら、特に最初の3つを検討対象にする。Gumbel分布は数学的にはGEVの一種だが、本稿では「リスクマップ」等になって両者を分けて扱う。

GEVの累積分布関数は

$$F(x) = \exp[-\{1 - \frac{x - \beta}{\alpha}\}^{1/\kappa}], \quad (\kappa \neq 0)$$

$$F(x) = \exp\{-\exp(-\frac{x - \beta}{\alpha})\}, \quad (\kappa = 0) \quad (1)$$

である (Coles 2001)。 x は同一分布・独立・多数のデータにおける一定期間 (例えば年) ごとの最大値であり、 F は値が x 以下である確率を表す。 F と再現期間 T とは $T(x) = 1/(1-F(x))$ の関係がある。 α 、 β 、 κ はそれぞれ「尺度パラメーター」「位置パラメーター」「形状パラメーター」と呼ばれ、 $F(x)$ のグラフの幅と x 軸上の位置、および形を決めるものである。 $\kappa = 0$ の場合は Gumbel 分布になる。(1) の逆関数は

$$x(F) = \beta + \frac{\alpha}{\kappa} \{1 - (-\log F)^\kappa\}, \quad (\kappa \neq 0)$$

$$x(F) = \beta - \alpha \log(-\log F), \quad (\kappa = 0) \quad (2)$$

である。この x は、再現確率 F あるいは再現期間 $T = 1/(1-F)$ に対する再現期待値を表す。

SQRT-ET は、江藤ほか (1986) が Gumbel 分布よりも裾の長い (long tail)、すなわち極端に大きい値が出やすい分布として提案したものである。累積分布関数は

$$F(x) = \exp\{-a(1 + \sqrt{bx}) \exp(-\sqrt{bx})\}, \quad (3)$$

である。この逆関数 $x(F)$ は (3) を数値的に解いて求められる。なお、この分布関数は国外ではほとんど使われていないようである。

LP3とGNOの関数形は小林 (2006) や外山・水野 (2002)、三浦・水野 (2005) に掲載されている。

以上の5つの分布のうち、Gumbel分布とSQRT-

ETはパラメーターが2個 (それぞれ α と β 、および a と b) であるのに対し、他の3つはパラメーター3個 (GEVの場合、 α 、 β 、 κ) である。他にもパラメーター3個の分布がいくつかある (Hosking and Wallis 1997; 外山・水野 2002など) が、本稿ではそれらには立ち入らないでおく。

各分布のパラメーターの推定法は、それぞれいくつかあるが、L-momentsを使う方法 (Hosking and Wallis 1997; 外山・水野 2002) が簡便であり、HoskingのホームページでFORTRANプログラムが公開されている (<http://lib.stat.cmu.edu/general/lmoments>; 2011年4月25日閲覧)。ただしSQRT-ETのプログラムはないので、以下の解析では小林 (2006) にならって最尤法を使った。

2.2 モンテカルロ・シミュレーションによる推定誤差の評価

年極値がある確率分布 (以下「母集団の確率分布」と表記) に従うものとし、そこから抽出される N 年間の年極値データに何らかの確率分布を適用して再現期待値を推定するとする。この「 N 年間の年極値データ」は確率変数であり、従って再現期待値の推定値も確率変数である。第1図は推定値の確率分布を模式的に描いたものに他ならない。本節では、モンテカルロ法を使ってこの確率分布を数値的に求め、それに基づいて (1) 再現期待値のバイアス、すなわち計算値の平均値と真値 (母集団の値) の比、および (2) 再現期待値の推定幅、すなわち計算値の変動幅を評価する。以後の議論では、母集団の確率分布として Gumbel 分布と SQRT-ET の2つを考え、これに本節の最初に触れた5つの確率分布を適用して、バイアス (推定値の平均値と真値との差) と推定幅 (推定値のばらつき) を観察する。この場合、適用する確率分布を以下「試行分布」と呼ぶ。

モンテカルロ・シミュレーションは下記の手順によった。

① 母集団の確率分布に従う乱数を N 個発生させ、 N 年分の年極値の擬似データ一式 (x_1, \dots, x_N) を作る。乱数の発生は、一様乱数 F を作り、これを累積分布関数とする $x(F)$ を算出する方法による。

N については主に $N=30$ と $N=100$ の場合を扱った。これらはそれぞれ、2010年時点でのアメダスと、明治時代にできた気象官署のデータ年数にほぼ相当する。

② x_1, \dots, x_N をデータとして、試行分布のパラメー

ターを求める。これに基づいて各再現期間（50年、200年等）の再現期待値を計算する。

- ③ ①②の手順（ラン）を10000回行い、結果を集計する。なお10000回とは、計算結果の推定幅を精度良く評価するのに十分なように設定したものである。

母集団の確率分布としては、Gumbel分布とSQRT-ETを扱う。「リスクマップ」による106年間の統計（1901～2006年）では、国内51官署のうち31地点でGumbel分布、16地点でSQRT-ETが選択されており（選択方法は第6節参照、他にGEVが3地点、LP3が1地点）、この2つで大半の地点がカバーされる。分布パラメータは、Gumbel分布については $\alpha=35$ mm、 $\beta=100$ mmとし、SQRT-ETについては $a=120$ 、 $b=0.5$ mm⁻¹とした。これらは「リスクマップ」の対象になった51地点への適用結果を参考にして、日本の標準的な地点における年最大日降水量のパラメータとして設定したものであるが、パラメータ値は以下の議論にとって本質的ではない。

3. 再現期待値のバイアスと推定幅

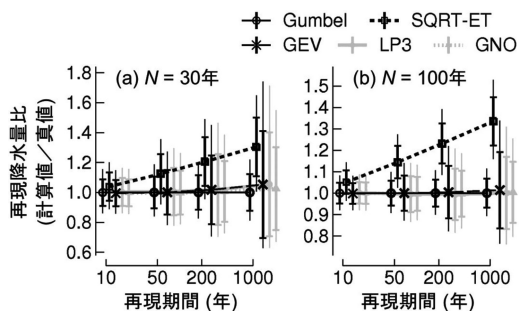
第2図は母集団がGumbel分布である場合について、横軸に再現期間を取り、縦軸には10000回のランによる再現期待値の平均値と標準偏差、および90%範囲（上・下位5%ずつ）を、いずれも真値（母集団の値）を1とした相対値で示したものである。試行分布を母集団と同じGumbel分布とした場合には、再現期待値の平均値はほぼ1であり、当然ながらバイアスはほとんどない。試行分布をSQRT-ETとすると、

再現期待値の平均値は1より大きく、正のバイアスが生ずる（第2図は再現期間10年以上の値だけ示しているが、もっと短い再現期間に対しては負バイアスになる）。これはSQRT-ETのlong tail性によるもので、SQRT-ETによるバイアスは再現期間が長くなるにつれて増す。一方、試行分布をGEV、LP3、GNOとすると、バイアスはほとんどなく、50年再現期待値（以下 P_{50} ）のバイアスは0.01以内、200年再現期待値（ P_{200} ）でも0.02以内にとどまる。

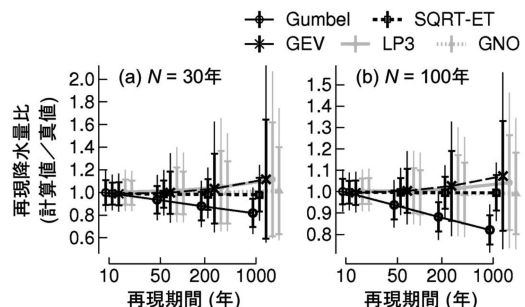
推定幅は、どの分布を使ってもある程度の大きさを持つ。試行分布がGumbel分布である場合、 P_{50} の90%範囲は $N=100$ なら0.91～1.10、 $N=30$ のときは0.84～1.18である。試行分布をSQRT-ETとした場合も、推定幅の大きさはこれらと同程度である。試行分布をGEV、LP3、GNOとしたときの推定幅は、10年再現期待値についてはGumbel分布やSQRT-ETを使ったときと同程度だが、再現期間が長くなるにつれて増加し、100年以上の再現期待値についてはGumbel分布やSQRT-ETを使ったときよりも明らかに大きくなる。

第3図は母集団がSQRT-ETである場合のものである。試行分布を母集団と同じSQRT-ETとした場合、バイアスはほとんどないが、第2図と同程度の推定幅があり、 P_{50} の90%範囲は $N=100$ なら0.89～1.11、 $N=30$ のときは0.80～1.20である。一方、試行分布をGumbel分布とすると負バイアスが生ずる（短い再現期間に対しては正バイアス）。GEV、LP3、GNOを当てはめたときの結果は第2図と同様であり、バイアスは小さいが、推定幅はGumbel分布やSQRT-ETを当てはめたときと同程度かそれ以上である。

上記のように、パラメータが2個であるGumbel



第2図 Gumbel分布を母集団とするデータに各試行分布を当てはめて計算した再現期待値比（計算値÷真値）の平均（記号）、標準偏差（太い縦棒）および90%幅（細い縦棒）。



第3図 第2図と同じ、ただし母集団はSQRT-ET。

分布や SQRT-ET を試行分布とすると、それが母集団の分布と異なる場合にはバイアスが生ずる。一方、パラメーター 3 個の各分布は母集団の関数形がどちらであってもバイアスは小さく、その一方で推定幅は大きい傾向がある。これは、推定の精密さと安定性（今の場合、バイアスと推定幅）とが相反する関係にあるという、統計解析の一般的事情に他ならない。問題は、この推定幅の大きさである。Gumbel 分布に従う母集団に Gumbel 分布を、あるいは SQRT-ET に従う母集団に SQRT-ET を適用するという、母集団と同じ分布を当てはめた場合でも、 P_{50} の推定値の 90% 幅は $N=100$ なら ± 0.1 程度、 $N=30$ のときは ± 0.2 弱の大きさを持つ。今回設定した母集団の場合、 P_{50} は 250 mm ぐらいなので、90% 幅はそれぞれ ± 25 mm 程度と $\pm 40 \sim 50$ mm になる。これは、正しい分布を適用してもなお、データが含む確率変動によって避けがたく生ずる推定誤差である。また、パラメーター 3 個の分布を当てはめた場合には推定幅がバイアスをはるかに上回り、従って、これらの関数のどれを使うかは実用上問題ではなくなる。

なお、今回の母集団として与えた SQRT-ET ($a=120$, $b=0.5 \text{ mm}^{-1}$) を GEV で近似すると、各パラメーターは $\alpha=31.3 \text{ mm}$, $\beta=94.0 \text{ mm}$, $\kappa=-0.123$ になる。この場合、SQRT-ET を GEV に置き換えても P_{50} は 0.3 mm, P_{200} も 5 mm の変化にとどまる。「リスクマップ」の対象になった 51 地点の 1901~2009 年のデータに SQRT-ET を当てはめてみると、 a は 17~442 の範囲に入る。これらも上記と同様、それぞれ $\kappa=-0.171 \sim -0.104$ の GEV で良く近似できる。

4. 適合度の評価方法と有効性

4.1 適合度基準

4.1.1 SLSC

SLSC の概念は、データを確率紙にプロットして関数を当てはめたときの、データ値と関数値の差（それぞれ標準化した値）を 2 乗平均したものである。式で表すと

$$\text{SLSC} = \frac{\sqrt{\sum_{i=1}^N (s_i - s_i^*)^2}}{N |s_{0.99} - s_{0.01}|} \quad (4)$$

ここで s_i^* は N 年間のデータのうち大きい方から i 番目の値、 s_i はその関数値に対応する「標準変量」であり、 $s_{0.99}$, $s_{0.01}$ はそれぞれ関数の 99% 値、1% 値に対

応する標準変量である。それらの定義は、

$$\begin{aligned} s_i &= -\log\{-\log F(x_i)\}, \\ s_i^* &= -\log[-\log\{(i-0.4)/(N+0.2)\}], \\ s_{0.99} &= -\log(-\log 0.99), \quad (s_{0.01} \text{ も同様}) \end{aligned}$$

である。このうち s_i^* の式はカナンプロットによる（小林 2006）。カナンプロット以外の定義のしかたもあるが、以下の議論にとって本質的ではない。

高棹ほか (1986) や宝・高棹 (1988) は「SLSC = 0.02 であれば十分な適合性を示しているといえる。SLSC > 0.03 であれば、他の分布形へのあてはめを試みた方がよい」としたが、「手引き」や「リスクマップ」では SLSC = 0.04 が適合のしきい値として使われている。

4.1.2 AIC

最小 2 乗法などによる関数当てはめの適合基準として、下記の AIC (Akaike's information criterion) が提案されている（赤池 1976）。

$$\text{AIC} = 2 \times (\text{自由パラメーター数} - \text{最大対数尤度})$$

今回の場合、

$$\text{最大対数尤度} = \sum_{i=1}^N \log f(x_i) \quad (5)$$

であり ($f = dF/dx$ は確率密度関数)、自由パラメーター数は Gumbel 分布と SQRT-ET は 2、GEV は 3 である。AIC の値が小さいほど適合性が良いと見なされる。

第 6 節では、SLSC などによる適合度判定結果と、AIC によるものを比べた結果を紹介する。

4.2. SLSC の特性

まず、母集団が Gumbel 分布である場合を扱う。第 4 図 a は試行分布として Gumbel 分布、GEV、SQRT-ET を与えた場合について、それぞれ 10000 回のランによる SLSC の平均値を、データ年数 N の関数として示したものである。SLSC は N とともに減少する。このことは葛葉 (2010) の指摘を裏づける。また、Gumbel 分布や SQRT-ET に比べて GEV は SLSC の平均値が小さい。これは、Gumbel 分布と SQRT-ET がパラメーター 2 個であるのに対して GEV はパラメーターが 3 つあり、データへの適応性が高いことを反映する。

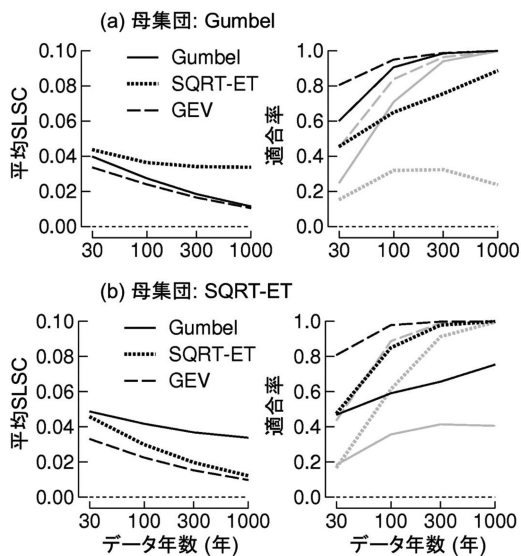
第4図bは、10000回のラン中にSLSCがしきい値以下になった割合（以下「適合率」と表記）と、 N との関係を示したものである。 $N=30$ のデータに対し、 $SLSC=0.04$ をしきい値とした場合、試行分布をGumbel分布としたときの適合率（すなわち正解率）は61%である。言い換えると、正しい分布（母集団と同じ分布）を与えても、39%の率で「適合しない」と判定される。その一方、SQRT-ETに対する「誤適合」が45%の率で発生する。適合率は N とともに増加し、 $N=100$ の場合にはGumbel分布による適合率（正解率）は90%に高まる一方、SQRT-ETによる誤適合率も65%になる。 $N=30$ の時よりも誤適合が増えるのは、 N が増すほどSLSCが小さくなり、 $SLSC=0.04$ のしきい値以下になることが増えるからである。

母集団がSQRT-ETである場合も同様の結果になる。 $N=30$ の場合、SQRT-ETを当てはめたときの正解率と、Gumbel分布の誤適合率はともに48%である。 $N=100$ の場合にはそれぞれ86%、60%になる。

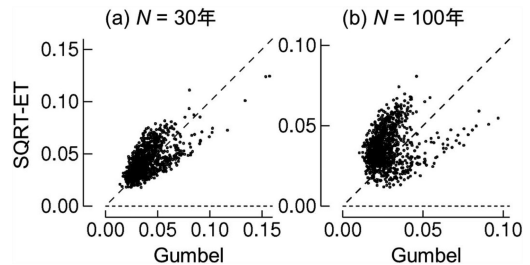
また、データ年数 N が小さいと、異なる試行分布によるSLSC同士に高い相関が存在する。第5図は、

Gumbel分布に従う母集団に対し、Gumbel分布を当てはめたときのSLSCとSQRT-ETを当てはめたときのSLSCの散布図である。 $N=30$ ならば、両者に0.64の相関がある。 $N=100$ になると相関係数は0.17に下がるが、散布図では $y=x$ の線（点線）をはさむ2方向に点が分布しており、その意味では2つのSLSC間に相関が見られる。もっと N が増すと、SLSCの値が全体に小さくなるとともに相関は弱くなり、 $N=300$ の場合にはほぼ無相関（相関係数=-0.05；図は省略）になる。

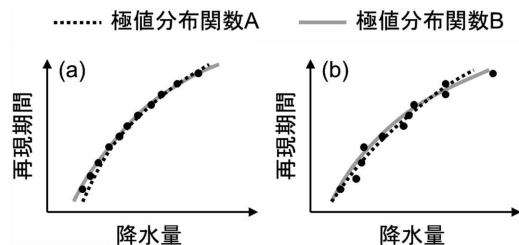
上記のことから分かるように、データ数が数十以下であるときには、2つの分布の一方に適合する／しないデータは、他方の分布にも適合する／しない傾向がある。これは、SLSCが母集団の関数形よりも、むしろデータのばらつき具合を反映することを示している。第6図はこのことを模式的に表したもので、確率紙上のデータに2つの関数を当てはめた状況を描いたものである。aはデータのばらつきが小さい場合であり、2つの関数ともそれなりにデータにフィットしている。一方、bはデータのばらつきが大きい場合であ



第4図 (a) Gumbel分布および(b) SQRT-ETを母集団とするデータに各試行分布を当てはめたときのSLSCの平均値(左)、および適合率(右; 濃い線は $SLSC=0.04$, 薄い線は $SLSC=0.03$ をしきい値としたもの)とデータ年数 N の関係。



第5図 Gumbel分布を母集団とするデータに、Gumbel分布とSQRT-ETをそれぞれ当てはめたときのSLSCの散布図。



第6図 データに2つの極値分布関数を当てはめた状態の模式図。(a)データのばらつきが小さい場合、(b)データのばらつきが大きい場合。

り、どちらの関数もデータへのフィットが悪くなっている。

5. ジャックナイフ法による再現期待値の変動幅評価

5.1 ジャックナイフ法の手順

ジャックナイフ法を使ってある統計量（例えば再現期待値；以下 P ）とその標準偏差を推定する手順は以下である。

- ① i 番目のデータを除いた $(N-1)$ 個のデータで確率分布関数のパラメーターを推定し、 P の値を求める。以下これを P_i と書く。
- ② ①を $i=1 \sim N$ のすべてについて行い、 P_i の平均値 \bar{P} と標準偏差 ΔP を求める。
- ③ N 個のデータすべてを使って計算された P の値を P_0 とすると、偏りを補正した P の推定値 P^* と、その標準偏差の推定値 ΔP^* は、

$$P^* = P_0 + (N-1)(P_0 - \bar{P})$$

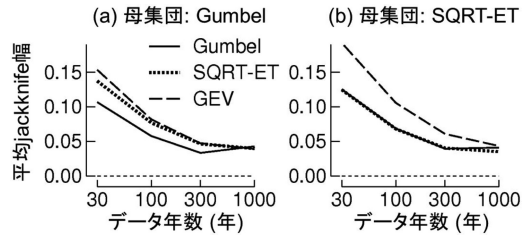
$$\Delta P^* = \sqrt{N-1} \Delta P \quad (6)$$

で与えられる（宝・高棹 1988）。以下、 ΔP^* を「ジャックナイフ幅」と表記する。

ジャックナイフ幅は、データの確率変動に伴う推定幅を与えるものであり、適合度とは別の尺度であるが、宝・高棹（1988）は分布関数の選択に当たり「 $SLSC < 0.03$ で同程度の適合を示したら、jackknife法を適用し、確率水文学の変動の小さいものを選ぶ」という方法を提案している。一方、「手引き」や「リスクマップ」では、Gumbel 分布、GEV、SQRT-ET のうち $SLSC \leq 0.04$ であるものを対象にして、ジャックナイフ幅が最も小さい分布を採用している。上記3つの分布の中に $SLSC \leq 0.04$ となる分布がないときは、LP3 と GNO を試行分布に加えて解析をやり直すことになっており、気象官署を対象にした106年間の統計では、全51地点の中で名古屋（LP3）だけがこれに該当する。ジャックナイフ幅を分布の選択に使うことの是非は第6節以降で議論することにし、本節ではジャックナイフ法の本来の趣旨である推定幅に関し、その特性を見る。

5.2 ジャックナイフ幅の特性

第7図は、母集団が Gumbel 分布である場合と SQRT-ET である場合について、 P_{50} のジャックナイフ幅（モンテカルロ・シミュレーションの10000回ラ

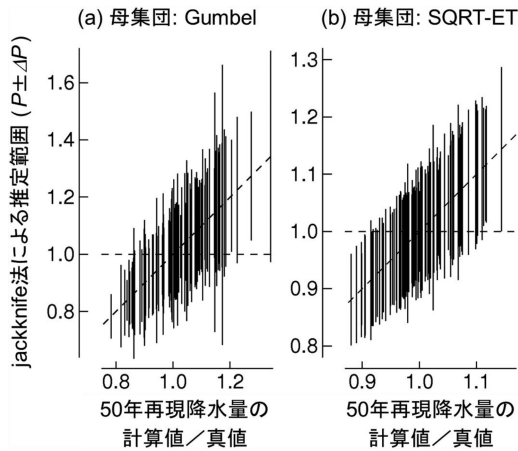


第7図 各試行分布を当てはめたときの、ジャックナイフ幅 ΔP^* の平均値とデータ年数 N の関係。

ンによる平均値) を、データ年数 N の関数として示したものである。母集団が Gumbel 分布であるときは、試行分布を Gumbel 分布としたときのジャックナイフ幅が最も小さい。母集団が SQRT-ET であるときには、試行分布を Gumbel 分布としたときと SQRT-ET としたときのジャックナイフ幅は同程度である。どちらの場合も、試行分布を GEV としたときのジャックナイフ幅が最も大きい。

母集団が SQRT-ET である場合には、当然ながら SQRT-ET を当てはめた方がバイアスが小さく、Gumbel 分布は第3節で見たように大きい再現期間に対する再現期待値を過小評価する。にもかかわらず、両分布を当てはめたときのジャックナイフ幅がほぼ等しいことは、関数の適合性（バイアスの小ささ）と安定性（推定幅の小ささ）とが別の概念であることを反映する。GEV のジャックナイフ幅が大きいのは、パラメーター数が多いためデータへの追従性が高く、それだけデータの確率変動に影響されやすいことによる。

次に、ジャックナイフ幅がデータの確率変動による推定幅の指標として適切であるかどうかを調べてみる。第8図 a は、母集団・試行分布をともに Gumbel 分布とし（言い換えると「正しい」分布を適用し）、 $N=100$ の場合について、横軸に各ランにおける P_{50} の計算値を取り、縦軸にジャックナイフ幅から算出した90%幅の目安を示したものである（図の見やすさを考え、10000回のうち100回のランのみ表示）。後者には、 P^* が自由度 $N-1$ の t 分布に従うことを仮定し、その90%幅（以下 t_{90} 幅）を充てた。ただしこれは、 P^* が t 分布に従うことを主張するものではなく、 t_{90} 幅が実用上使いやすい尺度であるという観点による便宜的な扱いであることを断っておきたい。もし t_{90} 幅が真値を含めば、 $y=1$ の線を横切るはずである



第8図 (a) Gumbel分布を母集団とするデータにGumbel分布を当てはめたときの、50年再現降水量(計算値÷真値)と、ジャックナイフ法で求めた P^* および ΔP^* の関係。(b) SQRT-ETを母集団とするデータにSQRT-ETを当てはめたときのもの。

ところ、実際には10000回のランのうちで88%がそうなる。第8図bは、母集団・試行分布をともにSQRT-ETとした場合の結果である。ここでは、全ランの83%で t_{90} 幅が真値を含む。

このように、 t_{90} 幅は真の90%幅をほぼ正しく表現する。しかし、わずかに t_{90} 幅が過小である。第8図を見ると、 P_{50} の計算値が真値よりも小さいラン(各図の左側に描かれたラン)で、 t_{90} 幅が $y=1$ の線に届かず、ジャックナイフ幅が過小評価される傾向がある。 P_{50} の計算値が真値よりも小さいランというのは、データの中に大きな値が(たまたま)少なく、その分データのばらつきが小さい場合に相当する。そのため、ジャックナイフ幅が過小になると考えられる。

なお、ジャックナイフ法に似た方法として、ブートストラップ法がある。これは5.1節の①「 i 番目のデータを除いた($N-1$)個のデータ」の代わりに、「 N 年間のデータから重複を許して抽出した N 個のデータ」を使うものである。これの性能評価も興味深いテーマではあるが、ジャックナイフ法でそれなりの成績が得られることを考え、今回はこの問題に立ち入らないでおく。

6. いろいろな適合度基準による判別結果比較

本節では、極値分布関数の選択方法として、以下の手法による結果を比較する。

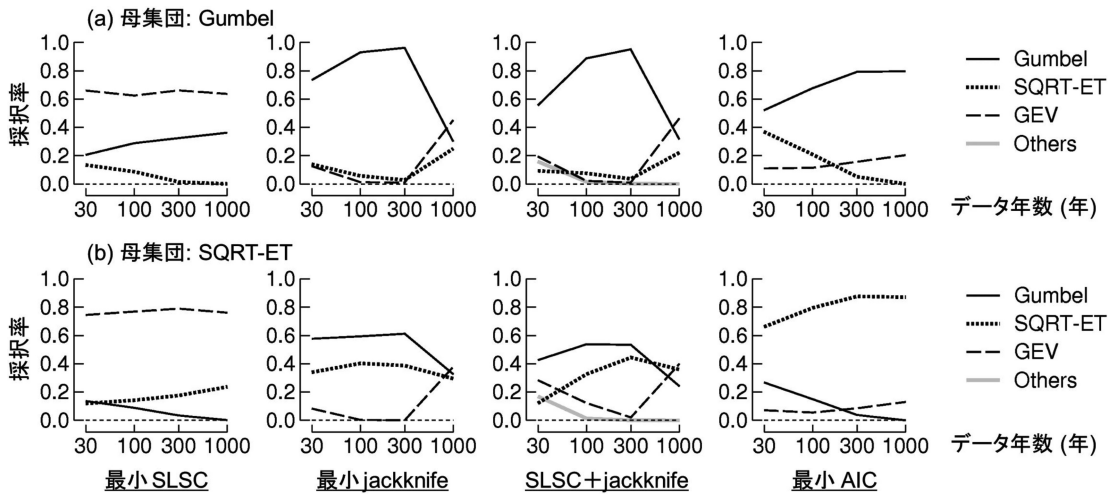
- SLSCが最小である分布(最小SLSC)。
- ジャックナイフ幅が最小である分布(最小ジャックナイフ幅)。
- $SLSC \leq 0.04$ である分布の中で、ジャックナイフ幅が最小である分布(SLSC+ジャックナイフ)。
- AICが最小である分布(最小AIC)。

このうち「リスクマップ」で使われているのは、「SLSC+ジャックナイフ」法である。

第9図は、母集団がGumbel分布とSQRT-ETである場合それぞれについて、上記の方法による各試行分布の採択率をデータ年数 N の関数として示したものである。最小SLSC法では、GEVの採択率が圧倒的に高い。これは、GEVがパラメーター3個の分布であり、データへの適応性が高いことを示している。

一方、最小ジャックナイフ幅を使うと、母集団がGumbel分布である場合だけでなくSQRT-ETである場合にも、Gumbel分布の採択率が最大になる($N=1000$ の場合は除く)。ジャックナイフ幅は推定幅、すなわち計算値の確率変動幅を与えるものであって、これは適合度(バイアスの大小)の尺度ではない。Gumbel分布のジャックナイフ幅が小さいのは、SQRT-ETに比べて裾が短い、すなわち極端に大きい値が出にくい性質により、推定値の変動が小さいことを反映する。

「SLSC+ジャックナイフ」法でも、最小ジャックナイフ幅を使ったときと同様、母集団がGumbel分布でもSQRT-ETでも、 $N=1000$ の場合を除いてGumbel分布の採択率が最大になる。母集団がSQRT-ETであってもGumbel分布が採択されやすいのは、 $SLSC \leq 0.03$ あるいは 0.04 を満たす分布に対して「ジャックナイフ幅最小」という条件を使った結果、多くのランにおいてジャックナイフ幅が決定力を持ち、適合度の良否に関わらず(すなわち、バイアスの有無に関わらず)推定幅の小さいGumbel分布が選択されやすいことを表している。5.2節で触れたように、宝・高棹(1988)は分布関数の選択に当たってジャックナイフ幅を考慮するよう提案はしたが、これには「 $SLSC < 0.03$ で同程度の適合を示したら」という注釈がついていた。言い換えると、宝・高棹(1988)の提案は、SLSCで甲乙つけ難ければ推定幅の小さい分布を採用しようという趣旨であり、「SLSC+ジャックナイフ」法とはジャックナイフ法の使い方が大きく違っている。



第9図 (a) Gumbel分布および (b) SQRT-ETを母集団とするデータにGumbel分布, SQRT-ETおよびGEVを当てはめたとき, 各判定基準で最適の分布と判定される率.

AICを使ったときは, 母集団がGumbel分布であればGumbel分布, SQRT-ETであればSQRT-ETの採択率がすべての場合に最大となる。これらの点で, AICはここで調べた方法の中では成績がいいと言える。

7. まとめと議論

(1) 再現期待値のバイアスと推定幅

極値統計解析でパラメーター2個の分布 (Gumbel分布, SQRT-ET) を使った場合, それらが母集団の分布と異なると, 再現期待値の推定結果にバイアスが生ずる。これに対してパラメーター3個の分布 (GEV, LP3, GNO) はいずれもバイアスが小さく, 50年再現期待値のバイアスは0.01以内, 200年再現期待値のバイアスも0.02以内にとどまる (実降水量にして数mm以内)。

一方, データの確率変動に伴う推定値の不確定幅 (推定幅) は, 母集団と同じ分布を当てはめてもなお無視できない大きさを持ち, 50年再現期待値の90%幅は $N=100$ なら ± 0.1 程度, $N=30$ のときは ± 0.2 弱である。また, パラメーター3個の分布による推定幅はパラメーター2個の分布を使うときと同程度かそれ以上であり, バイアスよりもはるかに大きい。

この結果と, Gumbel分布がGEVの一種であり, SQRT-ETもGEVで近似できることを併せ考えると, 実用上はGEVだけで事足りると言えよう。これは, データがGEVに適合することを主張するもので

はない。最初にしたように, GEVの理論的前提である「同一分布・独立・多数のデータ」という条件は実際には成り立つ保証がないのだが, だからと言って他の分布関数を使っても, データの確率変動による不確定さに阻まれて実用上意味のある改善にはつながらないというのがこの趣旨である。なお, この問題は二宮 (2011) のコメントに対する回答 (藤部 2011) でも触れたところである。

一方, GEVはパラメーター2個の分布 (Gumbel分布, SQRT-ET) に比べて推定幅が大きいという問題がある。藤部 (2010) で指摘されたように, GEVの形状パラメーター κ の推定誤差が再現期待値の精度を大きく左右する。そこで藤部 (2010) は, 再現期待値の推定幅を小さくするため, 地点ごとに求めた κ の値を領域平均し, κ を固定して計算しなおす方法 (その場合, 未知パラメーターは α と β の2個になる) を提案した。これは全域一律に同じ κ 値を与えるという単純なやり方であるが, もっと手の込んだ方法で地域区分を行い, 各地域内のデータを統合して処理する「地域頻度解析」も提案されている (Hosking and Wallis 1997; 外山・水野 2002)。また, 鈴木・菊地原 (1984) や石原 (2010) も何らかの形で複数地点のデータを組み合わせる方法を試みている。しかし, 母集団の地域的な均質・不均質をどう見極めるかが問題であり, この点はさらに検討を要する課題である。

(2) SLSC とジャックナイフ法の特徴

SLSC は、葛葉 (2010) が指摘したようにデータ年数に依存する。従って、SLSC による適合性判定のしきい値として定数 (0.03あるいは0.04) を使うことは適当でない。また、データの年数が小さい (数十年以下) 場合には、異なる試行分布による SLSC 同士に相関があり、SLSC は試行分布の母集団への適合性よりも、むしろデータのばらつき具合を反映する傾向がある。

ジャックナイフ法は、再現期待値の推定幅についてほぼ妥当な値を与える。しかし、推定幅の大きさと適合度 (バイアスの大小) とは別であり、十分な検討なしにジャックナイフ幅を分布関数の選択に使うと、かえってバイアスのある分布が選択される場合がある。従って、ジャックナイフ幅を適合性の判定に使うことは再考を要する。

最後に、いくつかの適合度判定法を比べた結果 (第9図) によると、AIC の成績が良かった。AIC は計算の簡便さの点でも優れている。AIC に似た情報量規準は他にもあり (統計数理研究所 1999; 下平ほか 2004など)、今後はこれらの規準についての検討が望まれる。

謝 辞

L-moments の計算には、Hosking によって公開されている Fortran77 のサブルーチン (2.1節) を利用した。

参 考 文 献

- 赤池弘次, 1976: 情報量規準 AIC とは何か. 数理科学, 14(3), 5-11.
- 中小河川計画検討会, 1999: 中小河川計画の手引き (案) ~洪水防御計画を中心として~. (財) 国土開発技術研究センター, 247pp.
- Coles, S., 2001: An Introduction to Statistical Modeling of Extreme Values. Springer, 208pp.
- 江藤剛治, 室田 明, 米谷恒春, 木下武雄, 1986: 大雨の頻度. 土木学会論文集, (369), 165-174.
- 藤部文昭, 2010: 極端な豪雨の再現期間推定精度に関する検討. 天気, 57, 449-461.
- 藤部文昭, 2011: 回答. 天気, 58, 147-151.
- 本城勇介, 篠田昌弘, 2008: 統計的手法による作用モデルの構築. 土木学会土木構造物荷重指針連合小委員会編「性能設計における土木構造物に対する作用の指針」, 土木学会, A-I-15-38.
- Hosking, J. R. M. and J. R. Wallis, 1997: Regional Frequency Analysis: An Approach based on L-moments. Cambridge Univ. Press, 224pp.
- 石原幸司, 2010: 都道府県別に適用した地域頻度解析によるアメダス地点の確率降水量評価. 水文・水資源学会誌, 23, 118-128.
- 小林健二, 2006: 確率雨量と再現期間の推定. 測候時報, 73, 51-72.
- 葛葉泰久, 2010: 治水計画策定における統計的手法—SLSC 及び費用便益分析に関する考察—. 土木学会論文集 B, 66, 66-75.
- 三浦大輔, 水野 量, 2005: L-moments を用いた強風の再現期待値の推定. 研究時報, 56, 97-125.
- 二宮光三, 2011: 「極端な豪雨の再現期間推定精度に関する検討」(藤部 2010) に対する質疑. 天気, 58, 143-145.
- 下平英寿, 伊藤秀一, 久保川達也, 竹内 啓, 2004: モデル選択: 予測・検定・推定の交差点. 岩波書店, 230pp.
- 鈴木昭夫, 菊地原英和, 1984: 異常豪雨を考慮した日降水量再現期間の推算法. 天気, 31, 179-189.
- 宝 馨, 高棹琢馬, 1988: 水文頻度解析における確率分布モデルの評価規準. 土木学会論文集, (393), 151-160.
- 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. 京都大学防災研究所年報, (29B-2), 157-171.
- 統計数理研究所, 1999: 特集「情報量規準」. 統計数理, 47, 273-424.
- 外山奈央子, 水野 量, 2002: L-moments を用いた地域頻度解析による全国アメダス地点における確率雨量の推定. 研究時報, 54, 55-100.

Discussion of Fitness Analysis for Selecting Distribution Functions in Extreme Value Analysis

Fumiaki FUJIBE*

* *Meteorological Research Institute, Tsukuba 305-0052, Japan.*
E-mail : ffujibe@mri-jma.go.jp

(Received 11 January 2011 ; Accepted 15 June 2011)

Abstract

A frequently used method in extreme value analysis is to select a distribution function that gives a best fit to data. To evaluate the performance of this approach, a series of Monte Carlo simulation was made by randomly sampling data from the Gumbel distribution or the square-root-exponential type maximum (SQRT-ET) distribution, and applying five candidate functions to calculate return values that were compared with exact values. It was found that two-parameter functions (Gumbel and SQRT-ET) cause biased results if they differ from the true distribution, while three-parameter functions (the generalized extreme value (GEV), log-Pearson type III (LP3), and generalized normal (GNO) distributions) have little bias that is much smaller than the range of uncertainty due to sample variability, indicating the unimportance of fitness analysis for these three-parameter distributions. Then the performance of the standard least-squares criterion (SLSC), which is widely used to quantify the fitness of extreme distribution functions to data, and also of the jackknife method, which is used for evaluating the range of uncertainty of return values, was examined by Monte Carlo simulation. It was found that the value of SLSC is dependent on data length, so that it is not appropriate to test the fitness using a constant threshold of SLSC. The jackknife method gives an almost correct value of the uncertainty range, but it is not suitable for fitness analysis because fitness and uncertainty are irrelevant to each other.