

水文頻度解析モデルの適合度の統計的仮説検定に関する考察

京都大学大学院工学研究科	学生員	○林 敬大
京都大学大学院工学研究科	正員	立川康人
京都大学大学院工学研究科	正員	萬 和明
京都大学大学院工学研究科	正員	Kim Sunmin
京都大学大学院工学研究科	正員	椎葉充晴

1 はじめに 我が国における河川計画の策定の際には、宝・高棹¹⁾によって提案された、SLSCによる適合度評価と Jackknife 法による安定性評価を併用したモデル選択手法が一般的に利用されている。しかしながら葛葉²⁾は、データのサンプル数が大きくなるにつれて SLSC の値は小さくなることを示し、SLSC によって適合度を評価する際にはサンプル数を考慮に入れる必要があることを指摘した。そもそも SLSC によって適合度を評価する際、0.03 ではなく 0.04 という値によって適合度の十分性が判断されることもあるようである。これは、田中・宝³⁾が河川流量に関しては 0.04 とすべきであると指摘したためであると考えられるが、宝・高棹¹⁾と田中・宝³⁾で扱われている水文資料のサンプル数に着目すると、前者が平均して 75 個程度であるのに対し、後者が平均して 42 個程度である。田中・宝³⁾は、河川水位から流量換算する際に生じるデータ特性を SLSC の値が比較的大きくなった理由として挙げているが、実はサンプル数が宝・高棹¹⁾のものよりも少なかったことがその主な理由として考えられるのである。

そこで本研究では、サンプル数に応じて実際にどのような SLSC の値で適合度を評価すればよいかを考察するために、統計的仮説検定の導入を試みる。

2 検定統計量としての SLSC

2.1 2 母数型モデルに対する SLSC の棄却域 水文データがある確率分布に従う母集団の実現値であると想定した上で、その確率分布と水文頻度解析モデルが一致するという仮説を帰無仮説、一致しないという仮説を対立仮説とする。そして SLSC を検定統計量として捉えなおし、有意水準 α に対応する SLSC の棄却域をデータの各サンプル数に応じて求め、第一種の過誤を起こす確率を異なるサンプル数に対しても統一的に評価することが本研究の目的である。そのためには帰無仮説のもとでの SLSC の確率分布を知る必要がある。

葛葉²⁾は帰無仮説を単純仮説としたときの正規分布に対する SLSC の確率分布を求めているが、本研究では帰無仮説を複合仮説とした場合の SLSC の確率分布を求める。例えば正規分布に対する SLSC の確率分布は下記の過程によって求める。

1. 正規分布 $N(\mu^*, \sigma^*)$ から N 個のサンプルデータ $x^j = (x_1^j, x_2^j, \dots, x_N^j)$ を発生させる。
2. サンプルデータ x^j の真の確率分布を正規分布であると正しく想定する。正規分布モデルのパラメータをある推定法によって求め、推定されたパラメータ $\hat{\mu}, \hat{\sigma}$ によってサンプルデータを標準化する。
3. SLSC の値 $S_{c,j}$ を求める (プロットングポジション公式は Cunnane 式を用いる)。
4. 1~3 の試行を 10000 回繰り返し、 $S_{c,j} (j = 1, 2, \dots, 10000)$ を得る。
5. 10000 個の $S_{c,j}$ の相対度数ヒストグラムを作成し、近似的に SLSC の確率密度関数を得る。

このとき求められる SLSC の確率分布は、データのサンプル数によって異なる (図 1)。そのため、SLSC の棄却域はサンプル数に応じて適切に変える必要がある。表 1 には各有意水準を設定した場合、実際にどのような棄却域によって適合度を評価すればよいかが示されている。また、各 SLSC 値に対応する超過確率も示されており、棄却域を固定する場合には第一種の過誤を起こす確率がサンプル数によって大きく異なることが分かる。

同様の手順によって、ゲンベル分布に対する SLSC の棄却域も容易に求めることができる。ただし図 2 が示すように、想定するモデルによって SLSC の検定統計量としての確率分布が異なるため、モデルによって棄却域が異なることに注意する必要がある。これは、SLSC の式中の順序統計量のばらつきがモデルの形状によって異なるためである。SLSC の値 S_c は次式で与

キーワード 水文頻度解析, SLSC, 統計的仮説検定, モンテカルロシミュレーション

連絡先 〒 615-8540 京都市西京区京都大学桂 C クラスター C1 棟, 電話: 075-383-3363

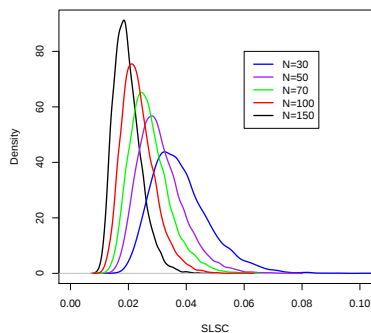


図1 正規分布・対数正規分布に対するSLSCの確率密度関数(複合仮説, 最尤法)

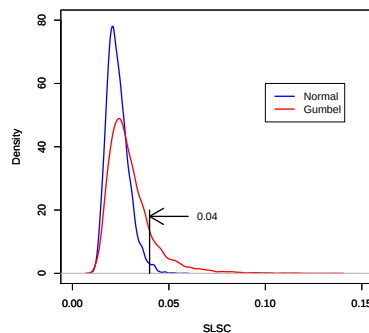


図2 正規分布, ゲンベル分布に対するSLSCの確率密度関数(サンプル数100, 複合仮説)

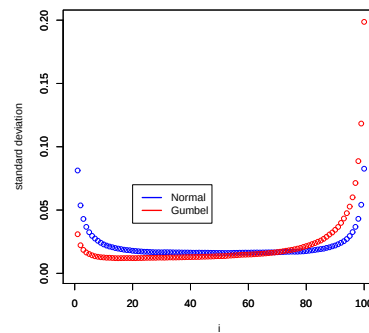


図3 $\frac{s(i)}{|s_{0.99} - s_{0.01}|}$ の標準偏差 ($s(i)$ は推定されたパラメータによって標準化された値)

表1 2母数型水文頻度解析モデルに対するSLSCの棄却限界値(複合仮説, 最尤法)

N	正規分布・対数正規分布						ゲンベル分布					
	有意水準 α			各SLSC値の超過確率			有意水準 α			各SLSC値の超過確率		
	0.10	0.05	0.01	0.02	0.03	0.04	0.10	0.05	0.01	0.02	0.03	0.04
30	0.0517	0.0567	0.0682	0.99	0.78	0.37	0.0681	0.0832	0.1188	0.99	0.82	0.51
50	0.0418	0.0461	0.0551	0.96	0.51	0.13	0.0575	0.0699	0.1010	0.97	0.67	0.33
70	0.0365	0.0402	0.0481	0.88	0.30	0.052	0.0507	0.0607	0.0861	0.92	0.52	0.23
100	0.0313	0.0344	0.0410	0.71	0.13	0.012	0.0446	0.0529	0.0756	0.84	0.39	0.15
150	0.0261	0.0284	0.0338	0.43	0.032	0.0012	0.0380	0.0453	0.0645	0.69	0.24	0.082

えられる。

$$S_c = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (s(i) - s_{(i)}^*)^2}}{|s_{0.99} - s_{0.01}|}$$

ここに, $s(i)$ はサンプルデータの標準変数, $s_{(i)}^*$ は $s(i)$ の期待値で, プロッティングポジション公式によって求められる。図3を見ればわかるように, ゲンベル分布のような右裾の長い分布は, その右裾部分で順序統計量のばらつきが大きくなる。結果として, 図2が示すように検定統計量としてのSLSCの値が大きくばらつくことになるのである。

2.2 3母数型モデルに対するSLSCの棄却域 水文頻度解析においてよく扱われる3母数型モデルにはGEV分布, ピアソン3型分布, 対数ピアソン3型分布が挙げられるが, これらのモデル分布は位置母数, 尺度母数に加えて, 形状母数を持つ。形状母数の値によってモデルの形状が変わり, 全く別のモデル分布としてふるまうため, 形状母数を自由母数とした複合仮説を導入することはできない。形状母数をあらかじめ固定母数として与えたモデルを考えれば, なおSLSCを検定統計量とした適合度の統計的仮説検定を行うことは可能となるが, 実際にどのような固定母数を与えればよいか, 客観的判断に窮する。また形状母数を自由母数として扱ったほうがモデルとして良くなる場合もあ

るにもかかわらず, その可能性を捨てることになる。

3 おわりに 第一種の過誤を異なるサンプル数に対して統一的に評価するためには, ある定められた有意水準に対するSLSCの棄却域によって適合度を判定すべきである。2母数型のモデルに対してはSLSCを検定統計量とした適合度の統計的仮説検定が容易に実装可能である。しかしながら, 検定統計量としてのSLSCの確率分布は想定するモデルの形状によって異なるため, 3母数型のモデルに対する, 形状母数を自由母数とする複合仮説を帰無仮説とした統計的仮説検定は, SLSCを検定統計量とした場合には困難となる。今後は, 情報量規準によるモデル選択手法についても考えていく予定である。

参考文献

- 1) 宝馨, 高棹琢馬: 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, No.393/II-9, pp.151-160, 1988.
- 2) 葛葉泰久: 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, Vol.66, No.1, pp.66-75, 2010.
- 3) 田中茂信, 宝馨: 河川流量の頻度解析における適合度と安定性の評価, 水工学論文集, Vol.43, pp.127-132, 1999.