

# 小標本への両側有界分布の適用

京都大学 防災研究所 正員 宝 馨  
 京都大学 防災研究所 正員 立川康人

P & G 土佐香織  
 京都大学 工学研究科 正員 市川 温

1 はじめに ある地点の 100 年確率降水量を推定する際、20 年や 30 年分の資料 (以後、小標本) しか手に入らない場合がある。そのような標本の観測期間が、たまたま少雨が続いた期間であれば、その地点に対して小さすぎる確率降水量を推定してしまう危険性がある。逆に、たまたま多雨が続いた期間であれば、その地点にとっては必要以上に大きな確率降水量を推定してしまうかもしれない (図 1 (a))。確率降水量を推定しようとしている地点に対して物理的合理性のある上限値を導入することにより、たとえ手元に小標本しかなく、またその分布が偏っている場合でも、データの蓄積が進んだ場合との誤差が比較的小さな値を得られる可能性が高まるのではないだろうか (図 1 (b))。本研究では、このような考えに基づき、水文頻度解析に上限値を導入することが、小標本の確率降水量の推定精度にどのような影響を与えるかを検証する。

EV4 分布 確率分布関数は以下のものである。

$$F(x) = \exp \left[ - \left\{ \frac{g-x}{\nu(x-a)} \right\}^\kappa \right] \quad (2)$$

母数は  $a$  (下限値),  $g$  (上限値),  $\nu, \kappa$ 。

さらに比較の対象として、上限無限大の 3 母数対数正規分布 (以下、LN3 分布) を用いる。

2 検証の手順 基本的には以下の方法で検証する。

2.1 擬似極値水文量時系列データの作成 母分布を想定し、その母分布に基づく乱数を 10000 個発生させ、擬似的な極値水文量時系列データ All とする。

2.2 小標本データセットの抽出 擬似極値水文量時系列データから、小標本データセットを抽出する。

2.3 確率水文量の推定 小標本データセットを用いて、確率水文量を推定する。適合度も求める。

2.4 評価 得られる 100 年確率水文量の推定値  $\hat{x}_{100}$  と推定精度を評価する。評価のために、次の 3 つの指標を用いる。

$$\text{相対誤差 RE} = \frac{\hat{x}_{100} - x_{100}}{x_{100}} \quad (3)$$

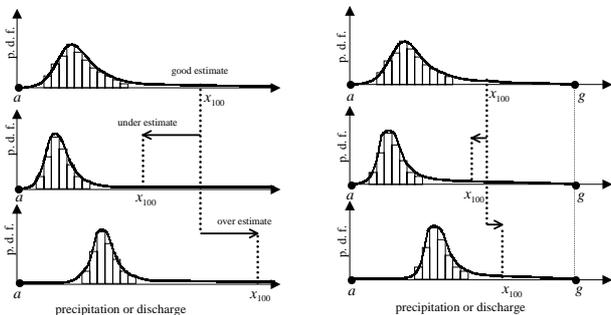
RE が 0 に近いほど良い推定結果が得られている。また、正の RE が得られたときは過大評価、負の RE が得られたときは過小評価であることを示す。

$$\text{平均相対誤差 MRE} = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{x}_{100} - x_{100})}{x_{100}} \quad (4)$$

$$\begin{aligned} \text{相対平方根平均平方誤差 RRMSE} \\ = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_{100} - x_{100})^2}}{x_{100}} \end{aligned} \quad (5)$$

RRMSE で真値からの絶対的なずれを、MRE で推定値の評価傾向 (過大評価か、過小評価か) がわかる。

3 一般的な小標本に対する特性評価 Slade 分布を母分布とする。発生させた All から、20 個または 30 個のデータから成る小標本 (以下 Size20, Size30 という) を、 $\{x_1, x_2, \dots, x_{20}\}, \{x_{21}, x_{22}, \dots, x_{40}\}, \dots$  というように抽出する。後は、2 で述べた手順に従って各分布の特性を評価する。Size20 と Size30 の場合に



(a) 下限値のみ (b) 上限値も導入  
 図1 上限値導入の効果

上下限値を持つ 2 つの分布モデルを用いる。このとき上限値は、世界記録及び日本記録に基づく可能最大 2 日降水量 (PMP)、さらに標本内最大値の 2 ~ 10 倍とする。また、下限値はゼロとする。

Slade 分布 確率密度関数は以下のものである。

$$f(x) = \frac{g-a}{(x-a)(g-x)\sigma_Y\sqrt{2\pi}} \times \exp \left[ -\frac{1}{2} \left\{ \frac{\ln \left\{ \frac{(x-a)/(g-x)}{\sigma_Y} \right\} - \mu_Y}{\sigma_Y} \right\}^2 \right] \quad (1)$$

母数は  $a$  (下限値),  $g$  (上限値),  $\mu_Y, \sigma_Y$ 。

表1 100年確率水文学推定値平均  $\hat{x}_{100}$  を用いた両側有界分布の有用性評価 [母分布: Slade 分布] 真値  $x_{100} = 312.37$  (カッコ内は RRMSE, 右側に MRE, \* は LN3 分布より優れているもの)

	$\hat{g}$	All	Size20	Size30			
Slade	$g_{world}$ (= 2652 mm)	313.25	0.003*	306.60 (0.138*)	-0.018	309.56 (0.114*)	-0.009*
	$g_{Japan}$ (= 1813 mm)	310.32	-0.007*	303.68 (0.135*)	-0.028	306.62 (0.111*)	-0.018
	$x_{max} \times 2$	301.97	-0.033	281.73 (0.154)	-0.098	286.29 (0.129)	-0.083
	$x_{max} \times 3$	307.83	-0.015*	292.22 (0.141*)	-0.065	296.22 (0.116*)	-0.052
	$x_{max} \times 4$	310.77	-0.005*	297.43 (0.139*)	-0.048	301.18 (0.114*)	-0.036
	$x_{max} \times 5$	312.53	0.001*	300.55 (0.138*)	-0.038	304.14 (0.113*)	-0.026
	$x_{max} \times 10$	316.05	0.012*	306.75 (0.141*)	-0.018	310.04 (0.116*)	-0.007*
EV4	$g_{world}$ (= 2652 mm)	560.02	0.793	514.87 (0.779)	0.648	528.01 (0.783)	0.690
	$g_{Japan}$ (= 1813 mm)	532.46	0.705	490.89 (0.682)	0.572	503.13 (0.688)	0.611
	$x_{max} \times 2$	465.06	0.489	363.66 (0.249)	0.164	378.31 (0.266)	0.211
	$x_{max} \times 3$	510.81	0.635	415.72 (0.415)	0.331	431.53 (0.439)	0.381
	$x_{max} \times 4$	536.54	0.718	446.92 (0.525)	0.431	463.19 (0.549)	0.483
	$x_{max} \times 5$	553.05	0.771	467.82 (0.602)	0.498	484.27 (0.624)	0.550
	$x_{max} \times 10$	588.74	0.885	516.11 (0.788)	0.652	532.43 (0.803)	0.704
LN3	-	319.55	0.023	312.89 (0.147)	0.002	315.88 (0.122)	0.011

表2 100年確率水文学推定値  $\hat{x}_{100}$  による両側有界分布の有用性評価 [母分布: Slade 分布] 真値  $x_{100} = 312.37$  (カッコ内は相対誤差 RE, \* は LN3 分布より優れているもの)

	$\hat{g}$	All (RE)	Min20 (RE)	Min30 (RE)	Max20 (RE)	Max30 (RE)
Slade	$g_{world}$ (= 2652 mm)	313.25 (0.003*)	247.87 (-0.206)	234.32 (-0.250)	379.24 (0.214*)	378.08 (0.210*)
	$g_{Japan}$ (= 1813 mm)	310.32 (-0.007*)	245.82 (-0.213)	233.30 (-0.253)	374.51 (0.199*)	372.94 (-0.194*)
	$x_{max} \times 2$	301.97 (-0.033)	225.00 (-0.280)	225.99 (-0.277)	341.29 (0.093*)	337.73 (0.081*)
	$x_{max} \times 3$	307.83 (-0.015*)	234.13 (-0.250)	229.28 (-0.266)	358.11 (0.146*)	355.38 (0.138*)
	$x_{max} \times 4$	310.77 (-0.005*)	238.69 (-0.236)	231.05 (-0.260)	366.19 (0.172*)	363.99 (0.165*)
	$x_{max} \times 5$	312.53 (0.001*)	241.43 (-0.227)	232.13 (-0.257)	370.93 (0.187*)	369.09 (0.182*)
	$x_{max} \times 10$	316.05 (0.012*)	246.87 (-0.210)	234.32 (-0.250)	380.22 (0.217*)	379.14 (0.214*)
EV4	$g_{world}$ (= 2652 mm)	560.02 (0.793)	459.11 (0.470)	385.63 (0.235*)	680.35 (1.178)	655.83 (1.100)
	$g_{Japan}$ (= 1813 mm)	532.46 (0.705)	440.29 (0.410)	374.32 (0.198*)	641.57 (1.054)	619.83 (0.984)
	$x_{max} \times 2$	465.06 (0.489)	304.58 (-0.025*)	302.32 (-0.032*)	452.03 (0.447)	442.30 (-0.416)
	$x_{max} \times 3$	510.81 (0.635)	354.24 (0.134*)	333.46 (0.068*)	533.42 (0.708)	518.87 (-0.661)
	$x_{max} \times 4$	536.54 (0.718)	384.18 (0.230)	350.84 (0.123*)	582.28 (0.864)	564.59 (-0.807)
	$x_{max} \times 5$	553.05 (0.771)	404.18 (0.294)	361.91 (0.159*)	614.85 (0.968)	594.97 (-0.905)
	$x_{max} \times 10$	588.74 (0.885)	449.77 (0.440)	385.70 (0.235*)	688.88 (1.205)	663.71 (-1.125)
LN3	-	319.55 (0.023)	252.29 (-0.192)	236.53 (-0.243)	389.24 (0.246)	389.01 (0.245)

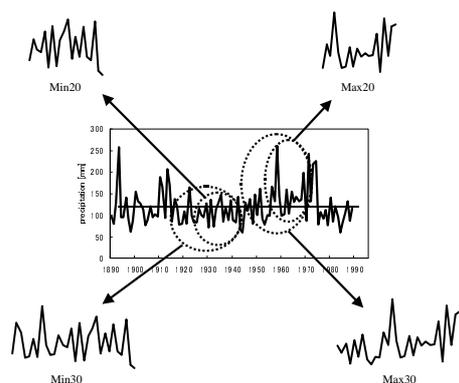


図2 データセットの抽出

得られた  $\hat{x}_{100}$  とその MRE を表 1 に示す。

4 特殊な小標本に対する特性評価 両側有界分布の持つ特性をさらに把握するため, 特殊な小標本を用いた評価を行う。そのため, 小標本データセットの抽出を 3 とは異なる方法で行う。その後は, 2 で述べた手順に従って行う。母分布は Slade 分布である。図 2 に抽出の様子を示す。

Min20, Min30 連続 20, 30 年分のデータだけを見たときに, 平均値が最も小さくなる部分から成るデータセット

Max20, Max30 連続 20, 30 年分のデータだけを見たときに, 平均値が最も大きくなる部分から成るデータセット

得られた  $\hat{x}_{100}$  とその RE を表 2 に示す。

5 結論 毎年最大水文学量を用いて確率水文学量を算定する場合, 小標本に確率分布を当てはめると, 求められる確率水文学量の値がかなり過大評価されたり過小評価されることが, しばしば経験されてきた。本研究では, モンテカルロ・シミュレーションによる検証を行うことにより, 近年提唱している両側有界分布を使用した場合に, どの程度の過大・過小評価が生じるのかを明らかにした。結果をまとめると以下のようなのである。

1. 一般的な小標本に対し, Slade 分布は 100 年確率降水量をやや過小評価するが, 推定精度は良くなる。このとき EV4 分布は過大評価する。
2. 小標本に含まれるデータが小さな値に偏っているとされる場合 (上記の Min20・Min30 のような場合) には, EV4 分布を用いることが過小評価の危険性を小さくするために有効である。
3. 小標本に含まれるデータが大きな値に偏っているとされる場合 (上記の Max20・Max30 のような場合) に Slade 分布を用いることは, 確率水文学量の過大評価を防ぐために有効である。

参考文献

[1] 土佐香織 (2000) : 両側有界分布の水文学頻度解析への適用に関する研究, 京都大学修士論文, pp. 51-77.