

リサンプリング手法による確率水文学の精度の評価について

京都大学大学院 学生員○清水 章
京都大学工学部 正 員 宝 馨
京都大学工学部 正 員 高 棹 琢 馬

1. 目的 水文学が従う確率分布を決定する際に適合度がその評価規準とされているが、同程度の適合度を示す確率分布が複数個ある場合、その中のどれを選べばよいのかということについては明らかにされていない。本研究ではJackknife 法、Bootstrap 法などのリサンプリング手法¹⁾による確率水文学の変動性を考慮に入れた確率分布の決定手法を提示する。さらに、標本数が確率水文学の推定誤差に及ぼす影響を、1912-1981年(70個)の琵琶湖流域における年最大k日降水量($k=1,2,3$)を用いて検討する。

2. Jackknife 法 ある確率分布Fに従う互いに独立なn個の確率変数を X_i 、その観測値を x_i (ただし、 $i=1, \dots, n$)、母集団の特性を表す量(平均値、分位値など)を θ 、その推定値(統計量)を $\hat{\theta}$ とする。いま、観測値 x_i ($i=1, \dots, n$)からi番目の値を除外した $n-1$ 個の標本から再び計算した統計量を $\hat{\theta}_{(i)}$ とする。Jackknife 法は $\hat{\theta}$ の偏倚を補正した推定値 $\hat{\theta}^*$ とその推定誤差 $\widehat{VAR}$ を次式のように与える。 $\hat{\theta}^* = n\hat{\theta} - (n-1)\hat{\theta}_{(.)}$, $\widehat{VAR} = (n-1)\sum_i(\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2/n$ (ただし、 $\hat{\theta}_{(.)} = \sum_i \hat{\theta}_{(i)}/n$)

3. Bootstrap 法 観測値 x_i ($i=1, \dots, n$)から反復を許して無作為にn個抽出し、その標本から統計量 $\hat{\theta}$ の推定値 $\hat{\theta}^{*1}$ を計算する。これを多数回(B回)繰り返し $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}$ を得る。Bootstrap 法は $\hat{\theta}$ の推定誤差 $\widehat{SD}$ を次式のように与える。

$$\widehat{SD} = \{ \sum_b (\hat{\theta}^{*b} - \hat{\theta}^{**})^2 / (B-1) \}^{1/2} \quad (\text{ただし、} \hat{\theta}^{**} = \sum_b \hat{\theta}^{*b} / B)$$

4. 確率分布の決定手法 図1は琵琶湖流域の年最大2日降水量を対数正規、Gumbel確率

紙にプロットしたものである。目視だけではどちらがよく適合しているのか判断できないことがわかる。また表1に適合度の評価規準の値を示したが(ただし、パラメタの推定は最尤法による)その値を比較するとSLSC¹⁾からは

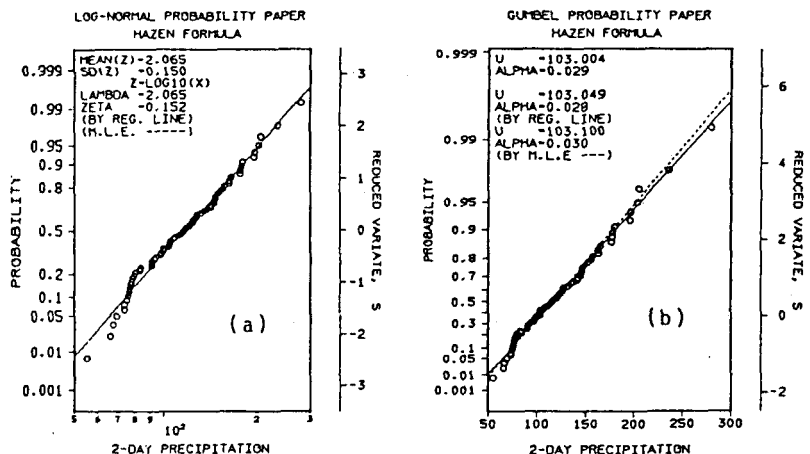


図1 Hazen 公式による琵琶湖流域の年最大2日降水量(70個)の対数正規確率紙 (a) とGumbel確率紙 (b)

Gumbel分布が、最大対数尤度(M.L.L.)・AICからは対数正規分布がよいという具合に異なる結論が導かれる。このような場合、確率水文学の変動性(推定誤差)に着目しその変動が小さな確率分布を選べばよいというのが本研究での提案である。表2の下段にこの2分布から求められる確率水文学の推定誤差を示した(ただしBootstrap法における抽出反復回数Bは1000回とした)。これらの値を比較すると確率水文学の変動が小さいGumbel分布がよいという結論が得られる。

表1 琵琶湖流域の年最大k日降水量における対数正規分布とGumbel分布の適合度の評価規準

k	確率分布	SLSC	M.L.L.	AIC
1	log-normal	0.02476*	-334.557	673.113
	Gumbel	0.02890	-334.387*	672.774*
2	log-normal	0.02498	-357.342*	718.683*
	Gumbel	0.02136*	-357.549	719.098
3	log-normal	0.01890*	-366.850*	737.701*
	Gumbel	0.02389	-367.031	738.061

*は適合度が良いことを示す

5.標本数と推定誤差の関係 Bootstrap法における抽出標本数n個をm個($m < n$)に置き換えた部分抽出法を用いて、標本数が確率水文学の推定誤差に及ぼす影響を検討する。表2の上段は年最大2日降水量について標本数の変化に伴う推定誤差の推移を示したものである。標本数が増加するにつれて推定誤差は減少する傾向を示し、その値は $n=70$ 個の標本による推定誤差に対して $m=50$ であれば約2割増、 $m=60$ であれば約1割増となることがわかる。また、どのmの値に対してもGumbel分布の推定誤差のほうが小さい。このことから、このデータに対してはGumbel分布のほうがよいと結論できる。

6.おわりに リサンプリング手法は確率分布の評価規準を与えたり、また確率水文学の推定に必要な標本数を評価する手段として有用であることを示した。今後、他の水文学に対してもこの手法の適用を試みてゆきたい。

表2 琵琶湖流域の年最大2日降水量による標本数mと確率水文学の推定誤差の推移
(反復回数1000回)

Methods	m	log-normal distribution				Gumbel distribution			
		30	50	100	200	30	50	100	200
Partial resampling with repetitions	20	219.2(27.8)	236.5(32.1)	260.0(38.2)	283.6(44.7)	214.9(24.7)	231.9(27.5)	254.8(31.3)	277.6(35.2)
	30	219.1(22.0)	236.4(25.4)	259.8(30.2)	283.3(35.3)	215.8(19.6)	232.9(21.8)	256.0(24.9)	279.1(28.0)
	40	218.5(19.2)	235.7(22.0)	259.0(26.2)	282.3(30.5)	215.9(17.1)	233.0(19.0)	256.2(21.7)	279.3(24.4)
	50	218.5(17.6)	235.8(20.3)	259.0(24.0)	282.3(28.0)	216.2(15.7)	233.4(17.5)	256.6(20.0)	279.7(22.4)
	60	218.3(15.7)	235.5(18.1)	258.8(21.4)	282.0(25.0)	216.1(13.9)	233.3(15.5)	256.5(17.7)	279.7(19.9)
Bootstrap	70	218.1(14.4)	235.2(16.5)	258.3(19.6)	281.5(22.8)	216.0(12.9)	233.3(14.4)	256.5(16.4)	279.7(18.4)
Jackknife	69	218.8(14.5)	236.1(16.7)	259.4(19.9)	282.8(23.2)	221.9(12.3)	239.6(13.8)	263.6(15.8)	287.4(17.9)
All samples	70	219.0	236.3	259.6	283.1	216.7	234.1	257.5	280.8

【参考文献】1)Efron,B.:The Jackknife,the Bootstrap and Other Resampling Plans, Regional conference series in applied mathematics,38.,Philadelphia,S.I.A.M.,1982
2)高棹・宝・清水:琵琶湖流域水文データの基礎的分析,京大防災年報第29号B-2,1986.