

降水量データに対する確率分布の適合度について

京都大学工学部 正 員 〇宝 馨

京都大学大学院 学生員 清水 章

京都大学工学部 正 員 高 棹 琢 馬

1. 緒言 水文頻度分析において確率分布モデルを選択する際に、データとモデルの適合度がまず検証される。適合度の悪いものはただちに最終モデルの候補から除外することができる。適合度は、“良いモデル”のための必要条件であり、モデル群をふるいにかける規準として有用である。ただし、適合度の評価規準はいくつかあり、良い評価を得るモデルが規準ごとに異なることが生じる。本研究では、2～3個の母数をもつ11種の確率分布を取り扱い、いくつかの降水量データに対するこれらの適合度を調べた結果を報告する。

2. データ 大阪の年最大日降水量(1889-1980年の92年間)¹⁾、大津、彦根の年最大k日降水量(1912-1985年の74年間)を対象とした。(k=1, 2, 3)

3. 確率分布モデル 正規分布(2母数)、対数正規分布(2母数, 3母数)、PearsonⅢ型分布(2母数, 3母数)、対数 PearsonⅢ型分布(2母数, 3母数)、平方根指数型最大値分布²⁾(2母数)、Gumbel分布(3母数)、対数Gumbel分布(2母数, 3母数)。

4. 母数推定法 すべての分布モデルに対して最尤法を用いることとし、改訂準ニュートン法による多変数関数の極小化プログラム DMINF1³⁾を利用する。上記11個のモデルをすべて最適化するのに要する計算時間は、大阪のデータの場合、京大大型計算機センターのFACOM M-382 で約6秒、M-780 で約3秒であった。

5. 適合度の評価規準 データとモデルを最尤法で適合させた場合、最大対数尤度が適合度の1つの評価規準となる。最大対数尤度は、 $MLL = \sum_i \log f(x_i, \theta)$ で与えられる。ここに、 f : 確率分布モデルの確率密度関数、 x_i : データ ($i=1, \dots, n$; n はデータの個数)、 θ : 母数 θ の最尤推定値である。一般に母数の個数が増えると適合度がよくなる。そこで、赤池⁴⁾はモデルの簡潔さ(母数の個数)をも考慮しうる情報量規準AIC(Akaike's information criterion)を提案した。

$AIC = -2 \log(\text{最大尤度}) + 2p = -2MLL + 2p$, p : パラメタ(母数)の個数。
われわれは、次に示すSLSC(standard least-squares criterion)を提案している⁵⁾。

$$SLSC = \sqrt{\xi^2} / |s_{0.99} - s_{0.01}|$$

ここに、 s_i : データの順序統計量 y_i に対応する標準変量($s_i = g(y_i, \theta)$)、 $s_{0.99}$, $s_{0.01}$: 非超過確率0.99, 0.01に対応する標準変量、 ξ^2 : 分位値 q_i に対応する標準変量 r_i と s_i との差の二乗平均。分位値は、 $q_i = (i - 1/2)/n$ (いわゆるHazen公式)で与えるのがよいとわれわれは考えている。SLSCは、標準変量にすべての母数を含めることのできないような分布にも適用可能である。標準変量に含まれない母数は最適化した値に固定して r_i と s_i を求めればよい。SLSCは、適合度の相対的評価規準としてだけ

Kaoru TAKARA, Akira SHIMIZU and Takuma TAKASAO

でなく絶対的評価規準としても有用である。SLSC=0.02であれば十分な適合を示す。また、SLSC>0.03であれば、他の分布を試みるべきである⁵⁾。すなわち、確率紙にデータをプロットしたり、データのヒストグラムと確率密度関数のグラフを描いたりすることなく、その値を計算するだけで適合度の良否が判定できるという利点をもつ。

6. 適合度の比較検討 大阪の年最大日降水量に対する各分布の適合度を表1に示した。大津、彦根および琵琶湖流域面積平均の年最大k日降水量の適合度をそれぞれ表2、3に示した。表の中で丸つき数字①②③は、各評価規準についての適合度の順位を表している。また、表の右半分には、30, 50, 100, 200年確率に相当する降水量を併記した。

【大阪の日降水量】3母数対数正規分布はどの評価規準についてもベスト3に入っており、

表1 大阪の年最大日降水量(1889-1980年の92年間) に対する適合度

		SLSC	M.L.L.	AIC	30	50	100	200
NORMAL	(2 PAR)	0.07937	-450.151	904.301	149.70	156.79	165.59	173.64
LOG-NORMAL	(3 PAR)	0.01666①	-432.818①	871.636②	164.02	179.70	201.71	224.69
LOG-NORMAL	(2 PAR)	0.02996	-434.914	873.828	153.92	165.09	180.07	194.97
PEARSON TYPE3	(3 PAR)	0.03765	-432.902②	871.803③	160.88	173.63	190.56	207.16
PEARSON TYPE3	(2 PAR)	0.06116	-438.172	880.344	151.27	160.49	172.42	183.81
LOG-PEARSON TYPE3	(3 PAR)	0.01749②	-432.910③	871.819	164.68	181.28	205.13	230.71
LOG-PEARSON TYPE3	(2 PAR)	0.10547	-463.170	930.341	282.30	332.33	408.57	495.63
SQRT-ET-MAX	(2 PAR)	0.02423	-433.091	870.182①	163.65	179.83	202.76	226.88
GUMBEL	(2 PAR)	0.04769	-434.414	872.829	153.10	164.77	180.50	196.17
LOG-GUMBEL	(3 PAR)	0.01859③	-433.167	872.335	164.82	182.07	207.14	234.35
LOG-GUMBEL	(2 PAR)	0.03496	-434.531	873.062	189.52	218.99	266.13	323.20

表2 大津の年最大k日降水量(1912-1985年の74年間) に対する適合度

		SLSC	M.L.L.	AIC	30	50	100	200	
k=1	NORMAL	(2 PAR)	0.05384	-364.832	733.663	167.16	174.52	183.65	192.01
	LOG-NORMAL	(3 PAR)	0.02120①	-357.207②	720.415	187.17	204.09	227.70	252.20
	LOG-NORMAL	(2 PAR)	0.02732	-358.297	720.593	176.14	188.31	204.57	220.69
	PEARSON TYPE3	(3 PAR)	0.02395②	-356.390①	718.780①	183.82	197.92	216.62	234.95
	PEARSON TYPE3	(2 PAR)	0.03205	-359.601	723.201	171.94	181.87	194.68	206.89
	LOG-PEARSON TYPE3	(3 PAR)	0.02489③	-357.466③	720.932	187.78	205.53	230.78	257.59
	LOG-PEARSON TYPE3	(2 PAR)	0.11386	-385.838	775.676	346.48	410.30	508.23	620.85
	SQRT-ET-MAX	(2 PAR)	0.03562	-357.668	719.336②	189.03	207.32	233.21	260.41
	GUMBEL	(2 PAR)	0.02666	-358.060	720.120③	176.98	190.22	208.07	225.85
LOG-GUMBEL	(3 PAR)	0.02882	-357.804	721.608	185.31	202.38	226.53	251.97	
LOG-GUMBEL	(2 PAR)	0.04661	-359.047	722.095	216.77	249.25	300.90	363.00	
k=2			SLSC	M.L.L.	AIC	30	50	100	200
	NORMAL	(2 PAR)	0.03634	-384.026	772.052	218.83	228.38	240.21	251.04
	LOG-NORMAL	(3 PAR)	0.02699①	-382.257③	770.514	231.87	246.77	266.39	285.50
	LOG-NORMAL	(2 PAR)	0.02882②	-382.480	768.959②	238.58	256.03	279.44	302.74
	PEARSON TYPE3	(3 PAR)	0.03137	-381.948①	769.896	232.39	246.99	265.96	284.17
	PEARSON TYPE3	(2 PAR)	0.02941③	-382.036②	768.072①	229.32	242.91	260.47	277.22
	LOG-PEARSON TYPE3	(3 PAR)							
	LOG-PEARSON TYPE3	(2 PAR)	0.11592	-410.378	824.756	484.70	579.58	726.78	898.11
	SQRT-ET-MAX	(2 PAR)	0.06684	-384.607	773.214	268.60	296.93	337.22	379.69
GUMBEL	(2 PAR)	0.04289	-382.838	769.676③	243.66	262.81	288.65	314.39	
LOG-GUMBEL	(3 PAR)	0.04403	-382.961	771.923	245.66	265.58	292.63	319.80	
LOG-GUMBEL	(2 PAR)	0.07514	-389.554	783.108	338.04	400.00	501.93	629.31	
k=3			SLSC	M.L.L.	AIC	30	50	100	200
	NORMAL	(2 PAR)	0.03710	-391.944	787.889	243.42	254.04	267.21	279.26
	LOG-NORMAL	(3 PAR)	0.01732①	-389.143⑤	784.287	257.37	274.07	296.11	317.66
	LOG-NORMAL	(2 PAR)	0.01994	-389.328	782.656②	263.33	282.29	307.71	332.97
	PEARSON TYPE3	(3 PAR)	0.01855②	-388.994①	783.988	257.16	273.19	294.00	313.97
	PEARSON TYPE3	(2 PAR)	0.01894③	-389.106②	782.213①	253.76	268.65	287.88	306.22
	LOG-PEARSON TYPE3	(3 PAR)							
	LOG-PEARSON TYPE3	(2 PAR)	0.12099	-420.103	844.205	508.14	608.44	764.29	945.99
	SQRT-ET-MAX	(2 PAR)	0.05437	-391.526	787.051	297.48	328.65	372.95	419.64
GUMBEL	(2 PAR)	0.02728	-389.616	783.232③	269.58	290.61	318.99	347.26	
LOG-GUMBEL	(3 PAR)	0.02912	-389.746	785.492	271.95	293.90	323.74	353.72	
LOG-GUMBEL	(2 PAR)	0.07079	-396.690	797.380	374.31	442.39	554.23	693.78	

SLSC, MLL についてはともに第①位, AIC については第②位である。その他の 3 母数の分布 (PearsonⅢ型, 対数 PearsonⅢ型, 対数Gumbel分布) はどれかの規準でベスト 3 に入っている。2 母数の分布では, 平方根指数型最大値分布はAICが第①位であり, SLSC, MLE についてはともに第 4 位である。SLSC=0.02423 であるから, まずまずの適合度である。他の 2 母数の分布はどれもベスト 4 にも入らない。ベスト 3 に 1 度でも入った 5 つの分布について確率水文量の値を見ると, 3 母数 PearsonⅢ型分布は他の 4 つの分布に比してかなり小さいめの値を与える。3 母数対数正規分布はリターンピリオドの大きいところでその他の 3 つの分布と比べてやや小さいめの値を与える。3 母数対数Gumbel分布は大きいめの値を与える。以下同様の検討を行う。

【大津の日降水量】 SLSC, MLL, AIC について, 3 母数 PearsonⅢ型分布がそれぞれ②①①位, 3 母数対数正規分布は①②④位であり, 適合性がよい。ただし, 3 母数PearsonⅢ型分布は, 3 母数対数正規分布や③③⑥位の 3 母数対数 PearsonⅢ型分布に比べて小さいめの確率水文量を与えることは大阪の場合と同様である。平方根指数型最大値分布は AIC が第②位であり, やや大きいめの確率水文量を与える。ただし, SLSC値が悪い。

【彦根の日降水量】 2 母数の分布でもよい適合性を示す。2 母数対数正規分布は②③①位であり, Gumbel分布, 平方根指数型最大値分布は MLL値が悪くないので AICではそれぞれ②位, ③位に浮上している。特に Gumbel 分布は SLSC 値もよい。3 母数対数正規分布,

表 3 彦根の年最大 k 日降水量 (1912-1985年の74年間) に対する適合度

			SLSC	M.L.L.	AIC	30	50	100	200
k=1	NORMAL	(2 PAR)	0.05238	-363.597	731.193	156.98	164.22	173.20	181.42
	LOG-NORMAL	(3 PAR)	0.01543①	-357.073②	720.146	170.34	184.10	202.82	221.71
	LOG-NORMAL	(2 PAR)	0.01603②	-357.171③	718.343①	167.54	180.15	197.13	214.06
	PEARSON TYPE3	(3 PAR)	0.01897③	-356.890①	719.781	169.14	181.48	197.72	213.50
	PEARSON TYPE3	(2 PAR)	0.02870	-358.101	720.203	162.30	172.30	185.24	197.61
	LOG-PEARSON TYPE3	(3 PAR)							
	LOG-PEARSON TYPE3	(2 PAR)	0.10245	-378.843	761.685	308.19	363.73	448.61	545.83
	SQRT-ET-MAX	(2 PAR)	0.03630	-357.776	719.553④	184.04	203.35	230.79	259.72
	GUMBEL	(2 PAR)	0.02015	-357.198	718.396②	168.33	181.59	199.48	217.30
	LOG-GUMBEL	(3 PAR)	0.01969	-357.190	720.380	170.56	184.77	204.19	223.82
	LOG-GUMBEL	(2 PAR)	0.05539	-361.265	726.530	226.04	266.20	331.90	413.48
k=2	NORMAL	(2 PAR)	0.06476	-384.554	773.107	204.23	213.84	225.75	236.66
	LOG-NORMAL	(3 PAR)	0.02071①	-376.259②	758.517	220.32	238.45	263.16	288.15
	LOG-NORMAL	(2 PAR)	0.02154②	-376.378③	756.755①	216.27	232.74	254.92	277.06
	PEARSON TYPE3	(3 PAR)	0.03391	-376.165①	758.330	219.29	235.65	257.19	278.16
	PEARSON TYPE3	(2 PAR)	0.04515	-377.592	759.183	209.79	222.88	239.82	256.01
	LOG-PEARSON TYPE3	(3 PAR)							
	LOG-PEARSON TYPE3	(2 PAR)	0.10758	-400.377	804.754	429.32	511.56	638.65	785.92
	SQRT-ET-MAX	(2 PAR)	0.03096	-376.968	757.937④	237.58	262.68	298.36	336.00
	GUMBEL	(2 PAR)	0.03089	-376.480	756.959②	217.38	234.62	257.87	281.04
	LOG-GUMBEL	(3 PAR)	0.02554③	-376.433	758.866	221.39	240.36	266.40	292.88
	LOG-GUMBEL	(2 PAR)	0.05479	-380.325	764.649	291.38	343.52	428.92	535.11
k=3	NORMAL	(2 PAR)	0.06146	-390.249	784.497	224.24	234.61	247.49	259.26
	LOG-NORMAL	(3 PAR)	0.01945①	-382.500①	771.000	241.58	260.98	287.35	313.97
	LOG-NORMAL	(2 PAR)	0.02047②	-382.621	769.243①	237.36	255.03	278.78	302.45
	PEARSON TYPE3	(3 PAR)	0.03061	-382.500①	771.000	239.63	256.90	279.60	301.65
	PEARSON TYPE3	(2 PAR)	0.04119	-383.783	771.567	230.50	244.61	262.86	280.30
	LOG-PEARSON TYPE3	(3 PAR)							
	LOG-PEARSON TYPE3	(2 PAR)	0.11196	-409.223	822.445	475.01	567.66	711.31	878.38
	SQRT-ET-MAX	(2 PAR)	0.03139	-383.186	770.372④	260.85	288.01	326.60	367.27
	GUMBEL	(2 PAR)	0.02745	-382.653	769.307②	239.00	257.72	282.98	308.15
	LOG-GUMBEL	(3 PAR)	0.02372④	-382.618③	771.237	242.73	263.06	290.89	319.10
	LOG-GUMBEL	(2 PAR)	0.05483	-386.646	777.291	319.23	375.26	466.70	579.97

3母数 PearsonⅢ型分布もよい適合を示している。(なお、対数 PearsonⅢ型分布については最適化がうまくいかない場合があった。プログラム改善の余地を残している。)

結局、地点日降水量に対しては、3母数対数正規分布、3母数 PearsonⅢ型分布がいつもよい適合を示す。2母数のものの中では、平方根指数型最大値分布がよい場合が多そうであるが、2母数対数正規分布やGumbel分布の方がよい適合を示すこともある。

【大津の2, 3日降水量】2母数 PearsonⅢ型分布が2日降水量、3日降水量のどちらに対しても③②①位でありよい適合を示す。他の2母数の分布のうちでは対数正規分布がよい。Gumbel分布はAIC値はよいが、SLSC値が悪い。2日降水量のデータをGumbel確率紙にプロットすれば直線上に乗らない。3母数対数正規分布もよい適合を示す。

【彦根の2, 3日降水量】大津と異なり、2母数 PearsonⅢ型分布の適合度は芳しくなく、かわりに2母数対数正規分布がよい。3母数の対数正規、PearsonⅢ型、対数Gumbelもよい。Gumbel分布、平方根指数型最大値分布はAICの値はよい(それぞれ②位、③位)が、SLSC値がよくない。

結局、地点2, 3日降水量については、2あるいは3母数対数正規分布がよい適合度を示す。ただし、大津では2母数 PearsonⅢ型分布の適合度がかなりよい。日降水量の場合と異なり、平方根指数型最大値分布はGumbel分布より適合度が悪い。平方根指数型最大値分布は右方に長く尾を引く分布であり、飛び抜けて大きな値の生じうる日降水量の場合にはよい適合度を示すことがあるが、2日以上になると降水量の値が相対的に平滑化される(右方に長く尾を引かなくてもよくなる)ので適合度が悪くなるものと考えられる。

7. 結語 年最大k日降水量に適合する分布の大体の傾向が把握できた。しかしながら、評価規準・適用例が限られていることもあり、さらなる検討を加えてゆく必要がある。適合度はモデルの screening に有用である。どの評価規準に対してもよい成績を示す分布モデルがあれば、それは有力な最終モデルの候補となる(モデル選択に携わる多くの人々を納得させうるという意味において)が、ここで見たように、あるデータセットに対して同程度の適合性を示す分布モデルが複数個存在する。こうした点に鑑み、われわれは、確率水文量の変動性を規準として確率分布モデルを評価する方法を提案している⁴⁾。

【参 考 文 献】

- 1) 神田・藤田：水文学－確率論的手法とその応用－，技報堂，1982，p.49.
- 2) 江藤・室田・米谷・木下：大雨の頻度，土木学会論文集，第369号/Ⅱ-5，1986，pp.165-174.
- 3) 富士通㈱：FACOM FORTRAN SSLⅡ使用手引書，1980，pp.403-406.
- 4) Akaike：A new look at the statistical model identification，IEEE Trans. Autom. Contr.，AC-19，1974，pp.716-723.
- 5) 高棹・宝・清水：琵琶湖流域水文データの基礎的分析，京大防災研年報，第29号B-2，1986，pp.157-171.
- 6) 宝・高棹・清水：水文統計解析における確率分布モデルの評価，京大防災研年報，第30号B-2，1987(投稿中)。