

II-4 情報と母数の数と最大エントロピー分布（その2）

信州大学工学部	正会員	寒川典昭
〃	〃	荒木正夫
矢作建設工業（株）	〃	船橋太道

1. はじめに

情報の与え方は、最大エントロピー分布のデータへの適合度を支配する重要な要因である。著者達^{1), 2)}は、昨年来、極値水文学を対象として、優れた情報の与え方を種々検討してきた。その結果、情報を与える関数 $g_r(x)$ ($r=1, 2, \dots, m$) として、巾関数と指数関数を用いた場合が良いことが明らかにされた。本稿では、千曲川流域5地点（長野、上田、松本、北牧、飯山）の年最大 k ($k=1, 2, 3$) 日降水量を用いて、最適な情報を与えるために、上述の両関数の具体的な形状と採用する母数の数を決定する。

2. 情報の与え方

$g_r(X)$ として、2母数 (MED(2p)) の場合は、 $x^s, \exp(-t \cdot x/M)$ (s, t : 正の整数, M : x の平均) の両者から1個ずつ、3母数 (MED(3p)) では、前者から2個、後者から1個、4母数 (MED(4p)) では、前者から3個、後者から1個採用する。これらの最大エントロピー分布は、次式ようになる。

$$\text{MED}(np): p(x) = \exp \left\{ -1 - \alpha - \sum_{i=1}^n \lambda_i x^{s_i} - \beta \exp(-t \cdot x/M) \right\} \quad (=(s_1, s_2, \dots, s_n, t)), \quad n=1, 2, 3 \quad (1)$$

ただし、 α, λ_i, β は分布のパラメタである。また、 s_i は情報とした統計モーメントの次数を表わしているため、分布形状との対応がつく4までの正の整数とし、 t もそれに準じることとする。

3. 推定分布の“良さ”の判定

s_i と t の最適値を決定するために、ここでは、宝・高樺・清水³⁾ が用いた LL, AIC, Jackknife 法、および、我々¹⁾ のリターンピリオドの安定性評価法を採用する。前2者はデータに対する全体的な適合度判定基準として、後2者は最大値水文学を対象とする時重要となる分布形の右裾の安定性の判定基準として有効である。すなわち、全体的な適合度に優れ、かつ、実用上、確率水文学あるいはリターンピリオドの安定した分布を最適な分布とみなすことにする。ただし、選定手順は宝らの方法による。

4. 実データへの適用結果

ここでは、1. で述べた15組の水文学へ適用する。ただし、2, 3母数の分布に焦点をあてて、最適な情報の与え方をみつけるものとする。4母数の分布は、母数の数を増加させることにより、更に改善の余地があるかを調べるために用いるものとし、松本、北牧、飯山の3地点への適用に留める。計算結果の掲載は、紙面の都合上、松本の年最大1日降水量の場合に限る。

Table 1 は、2母数、3母数分布の内、LL, AIC について、少なくともどちらか一方が、上位5位に入った分布の LL, AIC の値と、対象としたリターンピリオドに対応する Jackknife 法の結果を示したものである。LL では3母数の分布が、AIC では2母数の分布が上位にきている。次に、Jackknife 法の計算結果を見ると、どのリターンピリオドに対しても、(1, 4, 4) の推定誤差が小さくなっている。Fig. 1 は、Table 1 の分布の内、計算可能なものについて、時系列的データの前半分の最大値のリターンピリオドの安定性をみたものである。どの分布も変動が小さく優劣を付け難い。以上より、ここでは、Jackknife 法の推定誤差が最小であった (1, 4, 4) を最良の分布として選定する。Fig. 2 は選定された分布のヒストグラムへの適合を示したものである。なお、Jackknife 法が計算可能で、かつ LL が最上位の4母数分布 (2, 3, 4, 2) についても同様の計算を行った結果、LL, AIC, および Jackknife 推定誤差については、(1, 4, 4) より優れた値を示したが、リターンピリオドの安定性が計算できなかったため、候補から除外した。

Table 1 LL, AIC およびJackknife 法の結果（松本）

p (x)	L-L	AIC	Jackknife法によるT年確率水量推定値（推定誤差）(mm)			
			30年	50年	100年	200年
1 4	-316.280	638.561	115.85(6.67)	122.91(7.62)	133.09(9.07)	145.28(10.27)
2 4	-316.618	637.237	111.62(5.82)	118.77(6.41)	125.81(7.16)	134.40(7.86)
1 3	-316.741	637.481	111.33(6.26)	119.17(7.10)	126.45(8.46)	135.26(9.79)
1 4 4	-316.166	638.332	112.89(5.30)	119.86(5.67)	129.35(6.19)	139.01(7.07)
1 3 4	-316.222	638.444	113.41(5.47)	120.71(5.96)	131.38(6.78)	138.44(7.88)
1 2 4	-316.260	638.520	114.42(5.70)	121.66(6.34)	133.30(7.46)	143.99(8.97)
2 3 4	-316.560	639.119	111.25(5.89)	118.47(6.55)	126.39(7.34)	134.17(8.16)

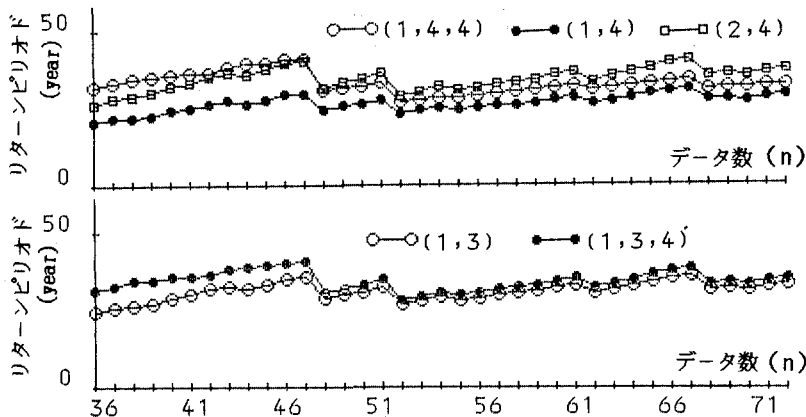


Fig.1 リターンピリオドの安定性（松本）

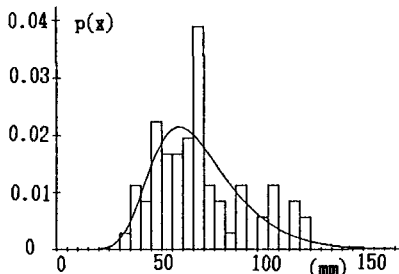


Fig.2 ヒストグラムへの適合（松本）

Table 2 最適な情報の与え方

	年最大k日降水量		
	1	2	3
長野	(1,3)	(1,4)	(1,3)
上田	(1,4,4)	(1,4)	(1,4)
松本	(1,4,4)	(1,3)	(1,4,4)
北牧	(4,4)	(3,4)	(2,3)
飯山	(1,4)	(1,4)	(1,2,4)

Table 2 は、以上の考察を他の14組の水水量について行って、選定した最良の分布である。全体的にみて、 $(t=1, s_i \text{の} 1 \sim 4)$ の分布が多く選ばれている。

5. あとがき

本稿では、ここで取り上げた水水量に対して、ほぼ最適な情報の与え方が明らかになった。今後は、何故 Table 2 の分布が最良であるかを、対象とした水水量のもつ特性との関連で検討していくとともに、ここでの議論を多変数に拡張したい。なお、本研究を行うにあたり、京都大学工学部、宝馨先生に、何かと御指導頂いた。計算には、信州大学大学院生、福沢直樹君の協力を得た。記して謝意を表する。

- 1) 寒川, 荒木, 大原: 情報の与え方と最大エントロピー分布（その2）, 第42回年講, II-3.
- 2) 荒木, 寒川, 船橋, 福沢: 情報と母数の数と最大エントロピー分布, 昭和62年度中部支部, II-51.
- 3) 宝, 高埜, 清水: 確率水水量の変動性を規準とした確率分布モデルの評価手順, 第42回年講, II-5.