

指数関数を情報に加えた2変数最大エントロピー分布(その2)

信州大学大学院 学生員○清水克彦
信州大学工学部 正会員 荒木正夫

信州大学工学部 正会員 寒川典昭
" 滝本信春

1. はじめに 最大エントロピー分布は、情報の与え方により様々な分布形を取り得ることが知られている。少ないデータから精度よく母集団を推定するには、適切な情報の与え方を模索する必要がある。我々¹⁾

²⁾は昨年度、従来の統計モーメントのみの情報に、指数関数の期待値も加えた2変数最大エントロピー分布(2MED)を提案し、その有効性を実データに対する分布形の全体的適合度の面から検討した。本稿は、その結果を踏まえ、リターンビリオドの安定性を見ることにより、2MEDの裾の部分の挙動を検討するとともに、相乗モーメントの高次化とパラメタ数増加の効果についても把握するものである。

2. 理論式 2MEDの一般式は、密度関数が具備すべき条件と任意関数 $g_r(\cdot)$ の期待値を制約条件とすると、

$$p(x, y) = \exp\{-\lambda_0 - \sum_{r=1}^N \lambda_r g_r(x, y)\} \quad (1)$$

と書ける。ここに、 λ_r ($r=1, 2, \dots, N$)はパラメタであり、 λ_0 は λ_r に依存して決定される。従って、パラメタ数は N 個である。いま、上式を $2M(g_1(\cdot), \dots, g_N(\cdot))$ (N :パラメタ数)と略記する。

3. 適用結果と考察 用いる実データは、立ヶ花地点の年最大流量に対応する小市、杭瀬下の32年間の最大流量であり、それらをそれぞれ確率変数 x, y とする。

3.1 分布形の裾の部分の検討 ここでは、図-1の手順に従い、 x, y ともに設定値を超過するリターンビリオドの安定性を見る。今回の設定値は、(a) x, y の平均の2.24倍(リターンビリオド100年付近)の値と、(b)立ヶ花1位流量に対応する x, y の値の、2ケースである。比較する分布は、統計モーメントのみを情報とする2パターンと、文献1)でLL, AICの上位5パターンである。表-1は採用した分布と計算結果であり、指数関数情報を加えた方の変動係数の値が(a)では0.11~0.16, (b)では0.13~0.21であるのに対し、統計モーメントのみの方は(a)では0.65, 0.74, (b)では1.45, 1.88と大きな値を示している。また、(b)のリターンビリオドの平均値を見ると、指数関数情報を加えた方は140~270年であるのに対し、統計モーメントのみの方は700, 820年とかなり大きくなっており、分布形が右裾の方にあまり尾を引いていないことがわかる。以上より、指数関数情報を加えた2MEDは、分布形の裾が安定し、既応最大流量に対しても妥当なリターンビリオドを与えると判断できる。

3.2 相乗モーメントの次数の検討 文献1)でLL, AICが最も良かった $2M(x, \exp(-3x/M_x), y^3, \exp(-y/M_y), xy)$ を用い、相乗モーメントの項を xy^2, x^2y, x^2y^2 に置き換え、その効果を検討した。表-2はそれぞれ分布のLL, AICの値を示したものである。ただし、 x^2y^2 のときはパラメタが同定できなかった。計算結果をみると、分布形の全体的適合度の面からは相乗モーメントの高次化の効果はなく、相乗モーメントの項は xy を用いた方がよいと判断できた。

3.3 パラメタ数の検討 文献1)でLL, AICの上位2位の分布を基として、統計モーメントによる情報を、 x, y に対しそれぞれ1つずつ増した7パラメタの2MEDと、基とした分布とを全体的適合度の面から比較した。表-3はその結果を示したものであり、パラメタ数が多い分だけ7パラメタの方がLLで

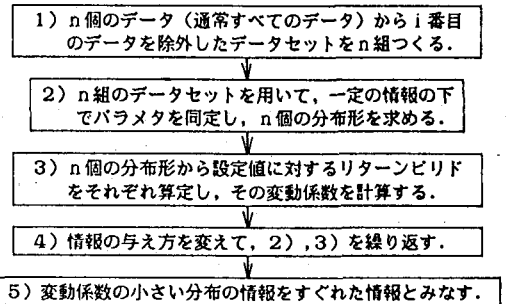


図-1 リターンビリオドの安定性の検討手順

よい値を示したが、パラメタ数を評価規準に取り込んだAICでは5パラメタの方がすぐれた分布と判断できた。すなわち、AICからみると統計モーメントの情報を追加することによる効果は期待できず、5パラメタで十分であると考えられる。

表-1 リターンビリオドの平均(年)と変動係数

p (x, y)	LL	AIC	設定値: 流量の平均値×2.24		設定値: 年最大流量1位	
			平均	変動係数	平均	変動係数
$2M(x, \exp(-3x/M_x), y^3, \exp(-y/M_y), xy)$	-510.0126	1030.025	101.497	0.129343	192.510	0.157170
$2M(x, \exp(-4x/M_x), y^3, \exp(-y/M_y), xy)$	-510.0738	1030.148	85.007	0.110083	148.764	0.131002
$2M(x, \exp(-2x/M_x), y^3, \exp(-y/M_y), xy)$	-510.2775	1030.555	121.034	0.160451	268.923	0.202658
$2M(x, \exp(-3x/M_x), y^3, \exp(-2y/M_y), xy)$	-510.3134	1030.627	93.572	0.153577	183.332	0.205522
$2M(x, \exp(-4x/M_x), y^3, \exp(-2y/M_y), xy)$	-510.3290	1030.658	80.101	0.131975	144.764	0.167088
$2M(x, x^2, x^3, y, y^2, xy)$	-513.8365	1039.673	119.730	0.646457	700.785	1.44832
$2M(x, x^2, y, y^2, xy)$	-514.0966	1038.193	118.531	0.739155	824.147	1.88369

表-2 相乗モーメントの高次化の効果

p (x, y)	LL	AIC
$2M(x, \exp(-3x/M_x), y^3, \exp(-y/M_y), xy)$	-510.0126	1030.025
$2M(x, \exp(-3x/M_x), y^3, \exp(-y/M_y), xy^2)$	-512.2181	1034.436
$2M(x, \exp(-3x/M_x), y^3, \exp(-y/M_y), x^2y)$	-512.9886	1035.977

表-3 パラメタ数の増加の効果

p (x, y)	パラメタ数	LL	AIC
$2M(x, \exp(-3x/M_x), y^3, \exp(-y/M_y), xy)$	5	-510.0126	1030.025
$2M(x, \exp(-4x/M_x), y^3, \exp(-y/M_y), xy)$		-510.0738	1030.148
$2M(x, x^2, \exp(-3x/M_x), y, y^3, \exp(-y/M_y), xy)$	7	-509.3631	1032.726
$2M(x, x^2, \exp(-3x/M_x), y^2, y^3, \exp(-y/M_y), xy)$		-509.3767	1032.753
$2M(x, x^3, \exp(-3x/M_x), y, y^3, \exp(-y/M_y), xy)$		-509.2925	1032.585
$2M(x, x^3, \exp(-3x/M_x), y^2, y^3, \exp(-y/M_y), xy)$		-509.3096	1032.619
$2M(x, x^2, \exp(-4x/M_x), y, y^3, \exp(-y/M_y), xy)$		-509.2157	1032.431
$2M(x, x^2, \exp(-4x/M_x), y^2, y^3, \exp(-y/M_y), xy)$		-509.2304	1032.461
$2M(x, x^3, \exp(-4x/M_x), y, y^3, \exp(-y/M_y), xy)$		-509.1550	1032.310
$2M(x, x^3, \exp(-4x/M_x), y^2, y^3, \exp(-y/M_y), xy)$		-509.1718	1032.344

4. おわりに 今回用いたデータセットでは、分布形の全体的な適合度と裾の部分の安定性の両面からみて、統計モーメントと指数関数の期待値を情報とした5パラメタの2MEDがすぐれていると判断できた。今後、さらに多くのデータセットへの適用を試み、対象とした水文量の特性と情報の与え方との関連性を検討していくとともに、本稿での議論を多変数、条件付きへと拡張していきたいと考えている。

1) 荒木・寒川・清水・能登谷：情報の与え方と2変数最大エントロピー分布，昭和62年度中部支部，II-52。

2) 清水・寒川・荒木：指数関数を情報に加えた2変数最大エントロピー分布，第43回年講，II-3。