II-3　　指数関数を情報に加えた２変数最大エントロピー分布

信州大学大学院　学生員　清水克彦
信州大学工学部　正会員　寒川典昭
〃　　〃　　荒木正夫

１．はじめに

最大エントロピー分布は，２変数の場合もまた，情報の与え方により様々な分布形を取ることが期待でき，少ないデータからうまく母集団を推定するためには，適切な情報の与え方を模索する必要がある．従来，我我[1]は，理論式は一般形で提案してきたものの，実データへの適用においては統計モーメントのみで情報を与え，データへの適合度を良くするためにパラメタ数を増加させてきた．しかし，そうして得られたパラメタ数の多い分布は，確率水文量の安定性，およびパラメタ同定に要する計算時間からみると，必ずしも最適な状態であるとは言えない．そこで本稿では，我々[2]が昨年度提案した統計モーメントと指数関数の期待値を情報とする，５個のパラメタをもつ２変数最大エントロピー分布を，１組の実データに適用し，従来の情報の与え方で９個のパラメタをもつ分布とデータに対する全体的な適合度を比較検討することにより，指数関数情報の有効性を把握する．

２．理論式

一般に２変数最大エントロピー分布は，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，次式のように書かれる．

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

従来のように，統計モーメントのみを情報とすると，（１）式は次式となり，2M(Na,Nb,Nc,Nd)と略記される．

$$p(x,y) = \exp\{-\alpha - \sum_{a=1}^{Na} \beta_a x^a - \sum_{b=1}^{Nb} \gamma_b y^b - \sum_{c=1}^{Nc}\sum_{d=1}^{Nd} \delta_{cd} x^c y^d\} \tag{2}$$

ここに，$\alpha, \beta_a, \gamma_b, \delta_{cd}$はパラメタであるが，$\alpha$は他のパラメタに依存して決定される．また，Na,Nbは最高４まで，Nc,Ndは２までとるものとする．つまり最高12パラメタまで取ることが可能である．一方，昨年度提案した２変数最大エントロピー分布は，パラメタ数を５個とし，指数関数も情報として加えることにより少ないパラメタ数で良い適合度を得ようとしたものである．その場合密度関数式は次のようになる．

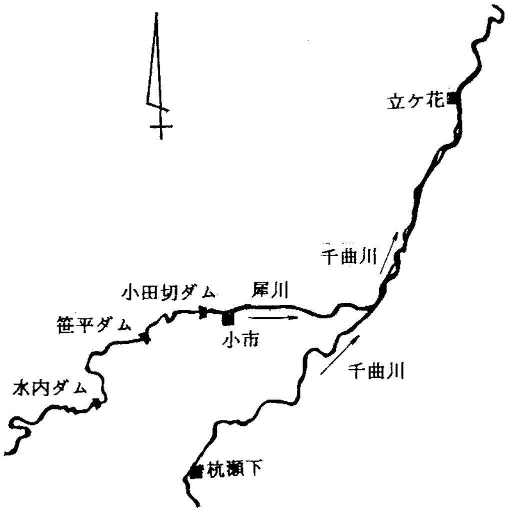
$$p(x,y) = \exp\{-\alpha - \beta_1 x^a - \beta_2 \exp(-bx/M_x) - \gamma_1 y^c - \gamma_2 \exp(-dy/M_y) - \delta xy\} \tag{3}$$

ただし，ａ，ｂ，ｃ，ｄは最大４の正の整数とする．

３．実データへの適用比較例

まずデータとして，図－１の位置関係にある，立ケ花地点の年最大流量に対応する小市，杭瀬下の32年間の最大流量を用い，それらをそれぞれ確率変数ｘ，ｙとして，（３）式より256パターンの最大エントロピー分布を求めた．表－１は上位10パターンの全体的な適合度評価値（Ｌ-Ｌ，ＡＩＣ）の値である．次に，従来の研究[1]で最も適合度が良いとされていた2M(4,4,1,1)を，最近の３年間のデータを除いた29年間のデータで求めた．これは，文献1)の研究が29年間のデータでなされていた事と，32年間のデータを用いると2M(4,4,1,1)が求まらなかったためである．したがって，表－１中の最上位の分布a=1,b=3,c=3,d=1についても，もう１度このデータ数で求め直した．図－２，図－３は，両分布の密度関数立体図である．また，表－２はそれぞれの分布のＬ-Ｌ，ＡＩＣの値である．これら２つの分布を比較すると，Ｌ-Ｌの値から見て，2M(4,4,1,1)の方がパラメタ数が多い分，実データによく適合していると言える．しかし，パラメタ数の多い分布は，水工計画上重要な裾の部分の安定性に問題を生じる恐れがあるとともに，パラメタ同定の計算時間が大きくなる．そこで，その数も考慮に入れた評価値ＡＩＣを見ると，a=1,b=3,c=3,d=1の分布の方が良い値を示している．す

なわち，このデータセットに対する適合度では，パラメタ数を５個とし，指数関数も情報として加えた場合の方が，簡便かつ良い適合度の分布形を得ることができると言えよう．



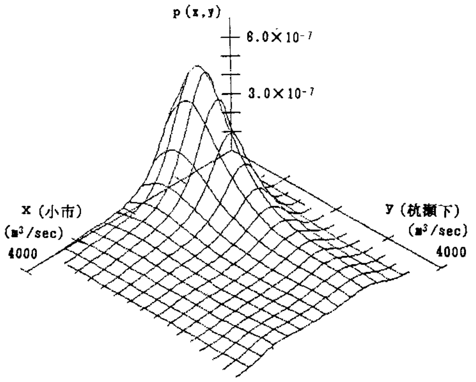
図－１　流量観測所の位置

表－１　上位10パターンのL-L，ＡＩＣ値

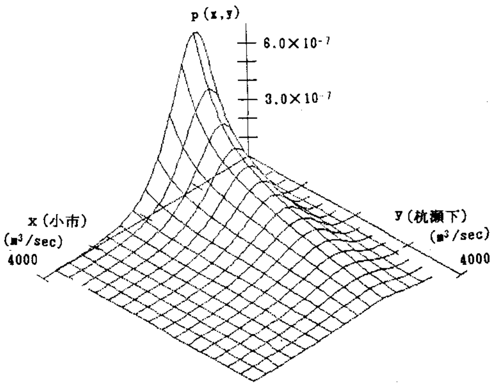
順位	分布の次数				L-L	AIC
	a	b	c	d		
1	1	3	3	1	-510.0126	1030.025
2	1	4	3	1	-510.0738	1030.148
3	1	2	3	1	-510.2775	1030.555
4	1	3	3	2	-510.3134	1030.627
5	1	4	3	2	-510.3290	1030.658
6	2	4	2	1	-510.4245	1030.849
7	2	4	2	2	-510.4326	1030.865
8	2	4	2	3	-510.6741	1031.348
9	1	2	3	2	-510.7048	1031.410
10	1	4	3	3	-510.8691	1031.738

表－２　L-L，ＡＩＣ値－データ数29個－

分布形	L-L	AIC
2M(4,4,1,1)	-456.3120	930.6239
a=1,b=3,c=3,d=1	-459.0062	928.0123



図－２　密度関数立体図
（９パラメタ，2M(4,4,1,1)）



図－３　密度関数立体図
（５パラメタ，a=1,b=3,c=3,d=1）

４．おわりに

本稿は１組のデータセットへの適用結果であるため，一般的なことはまだ言える段階ではないが，１変数の場合から類推すれば，統計モーメントと指数関数の期待値を組み合わせて情報とすると，極値水文量に対して良い適合度を示すとともに，安定した確率水文量が得られそうである．今後，確率水文量の安定性評価を含めて多くのデータセットへの適用を試みることにより，上述のことを検証するとともに，水文量のもつ特性と(a,b,c,d)の値との関係を明らかにしていきたい．なお，計算，作図等で，信州大学大学院生の能登谷敦君の協力を得た．記して謝意を表する．

１)寒川・荒木・寺島：２変数最大エントロピー分布の適用性，信州大学工学部紀要，第62号，PP.33-48，1987年９月．

２)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，土木学会中部支部昭和62年度研究発表会講演概要集，Ⅱ-52，PP.224-225，1988年３月．