

情報の与え方と2変数最大エントロピー分布

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
 信州大学大学院 学生員○清水克彦 信州大学工学部 能登谷敦

1. はじめに

著者達は、2変数最大エントロピー分布を一般形の記述で提案した¹⁾。しかし具体的にパラメータ同定する段階では、情報として統計モーメントのみに限ってきた。これは、統計モーメントが推定分布の形状と密接に関係していることから意義があるが、対象とする水文量ごとに適切な情報の与え方を検討する必要性が、1変数の場合²⁾から推察される。したがって、本稿では情報を与える関数 $g_r(\cdot)$ として何が良いかを1組の実データを用いて検討する。

2. 理論式

確率変数 x, y の同時確率密度関数を $p(x, y)$ とし、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、2変数最大エントロピー分布は次式で与えられる。

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

ここに、 λ_r はパラメータ(母数)であり、制約条件を満足するように決定される。式(1)は $g_r(\cdot)$ の与え方により様々な形状をとる。従来は、 $g_r(\cdot)$ として、巾関数のみで情報が与えられた。すなわち式(1)は、

$$p(x, y) = \exp\left\{-\alpha - \sum_{a=1}^{N_a} \beta_a x^a - \sum_{b=1}^{N_b} \gamma_b y^b - \sum_{c=1}^{N_c} \sum_{d=1}^{N_d} \delta_{cd} x^c y^d\right\} \tag{2}$$

となり、 $2M(N_a, N_b, N_c, N_d)$ と略記された。ここに、 $\alpha, \beta_a, \gamma_b, \delta_{cd}$ は λ_r と同様、パラメータである。今回は、 $g_r(\cdot)$ に巾関数と指数関数の両方を用いた。このとき式(1)は次式となる。

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

ここに、 M_x, M_y はそれぞれデータ x, y の平均値であり、相関は xy の期待値の形で取り込んでいる。

3. 実測データへの適用と考察

データは、Fig.1 に示す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、それぞれ確率変数 x, y として、2変数最大エントロピー分布を求めた。

いま式(3)の a, b, c, d を最大4とすると、109パターンでパラメータが同定できた。そこで得られた分布形の全体的な適合度の良否を LL(対数尤度), AIC により評価した³⁾。Table 1 は、上位5パターンの LL, AIC の値である。その結果、 a, d は低次で、 b, c は高次で情報を与えた場合のほうが良い値となった。また、そのとき x 軸断面図は指数関数的な分布を、 y 軸断面図は正規関数的な分布を示すことがわかった。Fig.2 は、109パターンのうち最も良い LL, AIC の値をもつ、 $a=1, b=3, c=3, d=1$ で情報を与えた場合の密度関数立体図である。Fig.3 は、従来の情報の与え方により、パラメータ数5個で実測データによく合っている $2M(2, 2, 1, 1)$ の立体図である。これを同じく LL, AIC を求め、前出の分布と比較した結果、今回の情報の与え方による分布形のほうが、ほとんどのパターンで良い値を示している。次に Fig.2, Fig.3 の分布と実測データのメッシュ確率を体積柱状図で示したものが Fig.4 である。64個のメッシュのうち37個の場合で、今回の分布の値のほうが実測データに近くっており、これからも適合度が改善されたと言える。以上のことから、パラメータ数5個の場合、情報として統計モーメントのみならず指数関数の期待値を含めたほうが、良い適合度が得られることがわかった。

Table 1 LL, AIC値

順位	分布の次数				LL	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

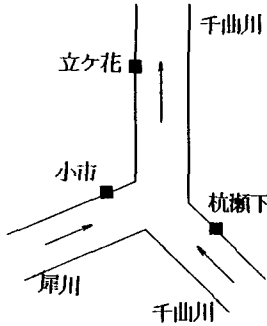


Fig.1 流量観測所の位置

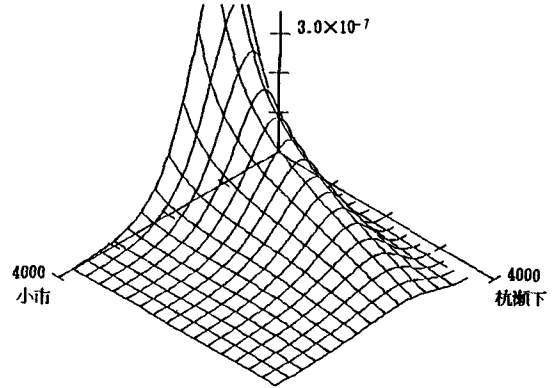


Fig.2 今回の2変数最大エントロピー分布
(a=1, b=3, c=3, d=1)

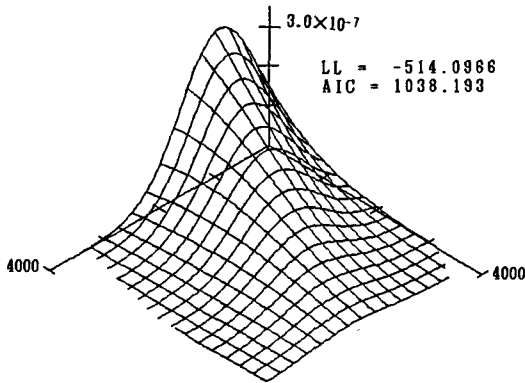


Fig.3 従来の2変数最大エントロピー分布
(2M(2,2,1,1))

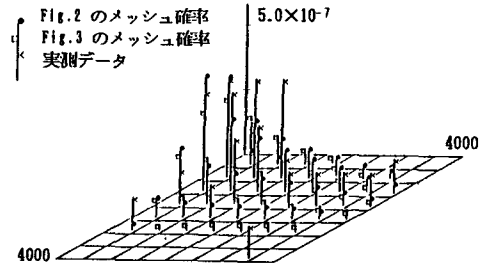


Fig.4 体積柱状図

4. おわりに

今回用いたデータの場合、指数関数の期待値も情報に加えたほうがデータへの全体的な適合度が改善されることがわかった。しかし、分布形の裾のほうの様子については、未解明である。また、他のデータの組については、まだ計算していない。今後、両者の検討をした上で、対象とする水文データごとに適切な情報を提示していきたい。

- 1) 寒川・荒木・寺島：2変数MEP分布とその特性に関する研究，第28回水理講演会論文集。
- 2) 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回土木学会年講，第2部。
- 3) 宝・高棹・清水：確率水文学の変動性を規準とした確率分布モデルの評価手順，第42回土木学会年講，第2部。