

任意関数の期待値を情報とした多変数 最大エントロピー分布とそれを用いた 千曲川合流地点の頻度分析

信州大学* 寒川 典昭
熊本工業大学** 荒木 正夫
福沢建設(株)*** 福沢 直樹

(1993 年 9 月 受付)

1. はじめに

相関のある 3 変数以上の水流量を同時に頻度分析されなければならない治水・利水計画問題が多々見受けられる。例えば、長沢 (1960) は、内水排除計画の場合、洪水の継続時間とその前後に降る雨、及びピーク水位が重要な水流量であり、こうした水流量の関連を統計的に処理することを要請している。また、長尾 (1975) は、河川堤防による治水計画では、洪水時のピーク水位、流量、流速等が破堤・溢水に関連する因子であり、それらを同時に確率変数と見なした統計解析が重要であることを述べている。石原・長尾 (1963) は淀川の合流河川の治水計画を合流前の 2 川の流量を 2 つの確率変数として処理しているが、合流前後の 3 川の堤内地が防災の対象となり、合流時差が大きい場合、合流後の流量も確率変数と見なさなければならず、3 変数問題として取り扱わねばならないことになる。

ところが、水文統計の分野では、それに対処する 3 変数以上の理論分布の方は、正規分布以外はほとんど利用されていないのが実状であった。数学的な興味においては、Krishnamoorthy and Parthasarathy (1951) の多変数ガンマ分布の研究があるが、水文統計に導入されるには解決されるべき問題がすくなくないようである。しかも、水流量が原型のまま正規分布に従うのは、単位時間の大きい、例えば年降水量のような特殊な場合に限られている。

このような状況から、水文学の分野において、寒川 他 (1986) は、Wragg and Dowson (1970) が情報理論の分野で展開した連続な $[0, \infty)$ 上の 1 変数最大エントロピー分布を多変数へと拡張した。こうして得られた連続な、それぞれの変数が $[0, \infty)$ 上の値をとる多変数最大エントロピー分布は、水流量の宿命ともいえる小標本性をうまく取り込み、利用において簡便で、水文データのヒストグラムに対する全体的な適合度がよく、しかも変数の数の上限に制限を持たないという特徴を有していた。ただし、ここでの多変数最大エントロピー分布には、Wragg and Dowson (1970) の場合と同様に、統計モーメントのみを情報としていたということ、及びデータヒストグラムへの適合度の客観的な評価並びに確率水流量やリターンピリオドの安定性の評価がなされていないということが今後の課題として残されていた。ここで、リターンピリ

* 工学部：〒 380 長野市若里 500.

** 〒 860 熊本市池田 4-22-1.

*** 〒 389-24 飯山市大字常盤 1234.

オドとは、水流量がランダムに生起するものとして、特定の大きさ以上の水流量の平均生起時間間隔を言う。複数の水流量を同時に取り扱う場合、これはそれぞれの水流量の特定の大きさを同時に超過する水流量の平均生起時間間隔となる。

本論文は上述の課題について、千曲川合流地点の治水計画問題を取り上げて、理論的に、及び実データを用いて検討したものである。得られた分布は、その密度関数式及びパラメータ同定法が簡便であり、3変数の水流量を同時に頻度分析するとき、データに対する全体的な適合度に優れ、更にリターンピリオドが安定するため、水文学的には興味深く、インパクトを与えるものと考えられる。従って、この分布は、土木工学の分野での河川の治水・利水計画の立案において、従来の便宜的な手法に代り、その計画の基本量の算定に利用されることが望まれる。ただし、第2章で記述しているこの分布の確率密度関数の推定、パラメータの同定法、及び統計学的な意味の解釈、更に第3章で記述しているこの分布のデータへの全体的な適合度の評価及びリターンピリオドの安定性の評価等に不備な点があるかも知れない。そのため、著者等は、これらの点について専門家の方々の統計学的な見地からの妥当性の検討を期待している。

2. 多変数最大エントロピー分布

2.1 確率密度関数の推定 (寒川 他 (1986))

n 個の連続な確率変数を $\mathbf{x}=(x_1, x_2, \dots, x_n)$, その確率密度関数を $p(x_1, x_2, \dots, x_n)$ とすると、多変数で定義された結合エントロピーは次式で与えられる。

$$(2.1) \quad H(x_1, x_2, \dots, x_n) = - \int \cdots \int p(x_1, x_2, \dots, x_n) \cdot \ln p(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

ここで、上式を次のように略記する。

$$(2.2) \quad H(\mathbf{x}) = - \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$$

以下では、上述の表現法に準じて理論式を記述する。

一方、確率密度関数の具備すべき条件と任意関数 $g_r(\mathbf{x})$ の期待値は

$$(2.3) \quad \int p(\mathbf{x}) d\mathbf{x} = 1$$

$$(2.4) \quad \int g_r(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} = E[g_r(\mathbf{x})], \quad r=1, 2, \dots, N$$

のように表される。ここに、 $E[\cdot]$ は期待値記号である。

次に、(2.3), (2.4) 式を制約条件として (2.1) 式を最大にする $p(\cdot)$ を求める。このことは、データ (母集団) から得られる情報は $g_r(\cdot)$ の期待値の形で取り入れ、それ以外は確率変数がランダムに発生する分布を算定していることになる。換言すれば、このことは情報 ($g_r(\cdot)$) の与え方には対象とする水流量に適した関数を採用しなければならないが、それ以外は最も客観的な分布を求めていることになる。著者らは水流量のように小標本で不確定さの強い現象の説明に、このような確率分布を導入することは妥当な立場と考える。

そこで、上述の問題をラグランジュの未定乗数法を用いて解く。ラグランジュ関数は次式で与えられる。

$$(2.5) \quad L = H(\mathbf{x}) + (\lambda_0 - 1) \left\{ 1 - \int p(\mathbf{x}) d\mathbf{x} \right\} + \sum_{r=1}^N \lambda_r \left\{ E[g_r(\mathbf{x})] - \int g_r(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \right\}$$

ここに、 λ_r ($r=1, 2, \dots, N$) はラグランジュ乗数であり、 $p(\cdot)$ から見るなら、これは推定分布のパラメータである。

(2.5) 式を $p(\cdot)$ について変分をとり、零とおくと

$$(2.6) \quad -\{\lambda_0 + \ln p(\mathbf{x})\} - \sum_{r=1}^N \lambda_r g_r(\mathbf{x}) = 0$$

となり

$$(2.7) \quad p(\mathbf{x}) = \exp\left\{-\lambda_0 - \sum_{r=1}^N \lambda_r g_r(\mathbf{x})\right\} \equiv M^{(n)}(g_1(\cdot), g_2(\cdot), \dots, g_N(\cdot))$$

が得られる。なお、 λ_0 は他のパラメータに依存して決定されるため、パラメータ数に入らない。(2.7) 式は多変数最大エントロピー分布の一般形であり、種々の $g_r(\cdot)$ を与えることにより、様々な形状を取り得る。

2.2 パラメータ同定法

ここでは、推定された確率密度関数が制約条件を満足するようにパラメータを同定する。

(2.7) 式を (2.3) 式に代入して整理すると

$$(2.8) \quad \exp(\lambda_0) = \int \exp\left\{-\sum_{r=1}^N \lambda_r g_r(\mathbf{x})\right\} d\mathbf{x}$$

が得られる。また (2.4) 式の右辺を簡略化して

$$(2.9) \quad E[g_r(\mathbf{x})] = \mu_r, \quad r=1, 2, \dots, N$$

と表し、(2.7) 式を (2.4) 式に代入して、(2.8), (2.9) 式を用いて整理すると、

$$(2.10) \quad \int g_s(\mathbf{x}) \exp\left\{-\sum_{r=1}^N \lambda_r g_r(\mathbf{x})\right\} d\mathbf{x} = \mu_s \int \exp\left\{-\sum_{r=1}^N \lambda_r g_r(\mathbf{x})\right\} d\mathbf{x}, \quad s=1, 2, \dots, N$$

となる。従って、(2.7) 式のパラメータ λ_r ($r=1, 2, \dots, N$) は、(2.10) 式 of 非線形方程式を解くことにより決定され、 λ_0 は (2.8) 式を変形した次式より求められる。

$$(2.11) \quad \lambda_0 = \ln\left[\int \exp\left\{-\sum_{r=1}^N \lambda_r g_r(\mathbf{x})\right\} d\mathbf{x}\right]$$

次に (2.10) 式を λ_r について解くために、Wragg and Dowson (1970) の方法を多変数に拡張して採用する。

まず、 λ_r の近似値として a_r を与え、残差を ε_r と置くと

$$(2.12) \quad \lambda_r = a_r + \varepsilon_r, \quad r=1, 2, \dots, N$$

となる。そこで、(2.12) 式を (2.10) 式に代入して、 ε_r に関してテイラー展開し、 $(\varepsilon_r)^2$ 以上の微小量の項を無視して整理すると、

$$(2.13) \quad \begin{aligned} & \sum_{t=1}^N \varepsilon_t \left[\int g_s(\mathbf{x}) g_t(\mathbf{x}) \exp\left\{-\sum_{r=1}^N a_r g_r(\mathbf{x})\right\} d\mathbf{x} \right. \\ & \quad \left. - \mu_s \int g_t(\mathbf{x}) \exp\left\{-\sum_{r=1}^N a_r g_r(\mathbf{x})\right\} d\mathbf{x} \right] \\ & = \int g_s(\mathbf{x}) \exp\left\{-\sum_{r=1}^N a_r g_r(\mathbf{x})\right\} d\mathbf{x} - \mu_s \int \exp\left\{-\sum_{r=1}^N a_r g_r(\mathbf{x})\right\} d\mathbf{x}, \\ & \hspace{15em} s=1, 2, \dots, N \end{aligned}$$

となる。ここで、 $g_0(\mathbf{x})=1$, $\mu_0=1$ とし、

$$(2.14) \quad c_{s,t} = \int g_s(\mathbf{x}) g_t(\mathbf{x}) \exp\left\{-\sum_{r=1}^N a_r g_r(\mathbf{x})\right\},$$

$$s=1, 2, \dots, N, \quad t=1, 2, \dots, N$$

とおくと, (2.13) 式は

$$(2.15) \quad \sum_{t=1}^N \varepsilon_t (c_{s,t} - \mu_s c_{0,t}) = c_{s,0} - \mu_s c_{0,0}, \quad s=1, 2, \dots, N$$

となる。さらに

$$(2.16) \quad c_{s,t} - \mu_s c_{0,t} = p_{s,t}, \quad s=1, 2, \dots, N, \quad t=1, 2, \dots, N$$

$$(2.17) \quad c_{s,0} - \mu_s c_{0,0} = q_s, \quad s=1, 2, \dots, N, \quad t=1, 2, \dots, N$$

とおくことにより, (2.15) 式から

$$(2.18) \quad \sum_{t=1}^N p_{s,t} \varepsilon_t = q_s, \quad s=1, 2, \dots, N$$

が得られる。上式は ε_t に関する N 元連立 1 次方程式であり, 方程式の数と未知数の数が一致するので解き得る。

3. 判定基準

水文頻度分析に用いられる分布の適合度の良否は, 未知の母集団から発生したデータとの比較においてなされる。あるデータセットに対して多数の推定分布が候補に挙がった場合, それらを比較検討して最良の分布を選択する必要がある。従来, その選択方法としては, 確率紙上へのプロットデータの直線性, 及びヒストグラムと分布曲線との適合性等から判断してきた。しかし, このような判断基準では, そのままでは多変数に対して適用することは難しく, また実際の水工計画で用いられる超過確率やリターンピリオドを得るのに必要な分布の裾の部分の安定性については無力であった。

そこで, 本稿では, 分布形の全体的な適合度に関しては L-L (Log-Likelihood) を用いる。また, 密度関数立体図とメッシュ確率図を利用して, 視覚的な判断により, 全体的な適合度比較の補助的手段とする。裾の部分の挙動に関しては, リターンピリオドの安定性評価法を採用する。以下に, これらの手法について概説する。

3.1 全体的適合度

(1) L-L

母集団が未知であるため, 推定された分布が複数ある場合, それぞれの L-L を計算し, その値の大きい分布が母集団に最も近い推定分布と判断する。一般に L-L は次式から算定される。

$$(3.1) \quad L-L = \sum_{i=1}^m \ln p(\mathbf{x} | \hat{\boldsymbol{\theta}})$$

ここに, $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{ni})$ ($i=1, 2, \dots, m$) は観測データであり, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$ はパラメータの推定値である。

(2) 密度関数立体図

3変数の分布形を直接グラフに描くのは不可能である。そのため, 3変数確率密度関数を2変数の周辺確率密度関数に変換して立体図を描き, 密度関数の形状を把握する。

(3) メッシュ確率図

密度関数立体図と同様に、周辺確率密度関数を求め、それを特定のメッシュに分割し、1つのメッシュ体積をメッシュ面積で除した値を表現したグラフであり、最大エントロピー分布とデータの特定のメッシュにおける適合度を見ることができる。

3.2 リターンピリオドの安定性

本論文での実データへの適用は、後述するように3変数に限っているため、ここでは2変数の場合に寒川 他 (1991) が行った1組のデータセットから一定の規則に従ってサンプリングを繰り返すリサンプリング手法を3変数に拡張する。具体的な手順は以下のようである。

- 1) m 個のデータ (通常はすべてのデータ) から i 番目のデータを除外したデータセットを m 組作成する。
- 2) m 組のデータセットを用いて、一定の情報の下でパラメータを同定し、 m 個の分布形を求める。
- 3) m 個の分布形からある設定値に対するリターンピリオドをそれぞれ算定し、その変動係数を計算する。
- 4) 情報の与え方を変えて、2), 3) を繰り返す。
- 5) 求めた変動係数を比較し、小さい値を示した分布を、裾の部分の安定した優れた分布と判断する。

なお、リターンピリオドは次式により計算する。

$$(3.2) \quad \frac{1}{1 - \int_0^{x_0} p(\mathbf{x}) d\mathbf{x}}$$

ここに、 x_0 はリターンピリオドを計算するための設定値である。

4. 千曲川合流地点への適用と考察

4.1 全体的な適合度

ここでは、図1に示した千曲川流域内の立ヶ花地点における年最大流量とそれに対応する小市、杭瀬下の最大流量を実データとして採用して、千曲川合流地点の治水計画問題を論じる。小市、杭瀬下、立ヶ花、それぞれの地点の流量データを確率変数、 x, y, z と表示する。なお、データの採用期間は1953~1985年(1954年欠測)の32年間分である。これらのデータは小市、杭瀬下の流量の平均値に大きな差はなく、立ヶ花地点の平均値は小市、杭瀬下の平均値の和に近い値を取っている。

いま、以上の資料と Sogawa et al. (1989) の $M^{(3)}(x^a, \exp(-bx/M_x), y^c, \exp(-dy/M_y), z^e, \exp(-fz/M_z), xy, yz, zx)$ ($a, b, c, d: 1\sim 4$ の整数, $M_x, M_y, M_z: x, y, z$ の標本平均) の z 項を省略して行った研究から優れた結果が得られている $a=1, b=4, c=3, d=1$ を固定してパラメータの同定を試みた結果、10種類の3変数最大エントロピー分布が同定された。ここで得られた推定分布のデータに対する全体的な適合度の良否をL-Lにより評価した。その結果を、上位5パターンの分布と $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ について表1に示しており、 $(e, f)=(1, 3)$ の分布が最良と判断された。また、更にここで求めた上位5パターンの分布と $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ のL-Lを比較すると、指数関数の期待値を情報に加えた上述の上位5パターンの分布の場合にかなり全体的な適合度の改善が読み取れた。

図2は $(e, f)=(1, 3)$ の分布の密度関数立体図である。この図について考察すると次のよう

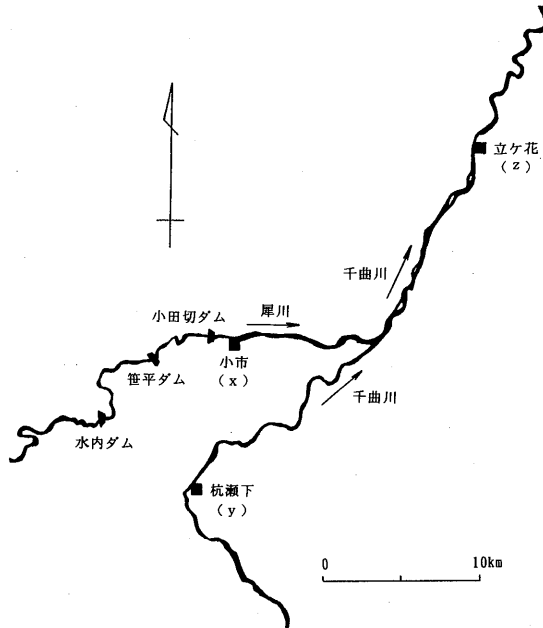


図1. 流量観測所の位置.

表1. L-Lの値.

分 布	L-L
$(e, f)=(1, 3)$	-0.77452×10^3
$(e, f)=(1, 4)$	-0.77480×10^3
$(e, f)=(1, 2)$	-0.77484×10^3
$(e, f)=(3, 4)$	-0.79461×10^3
$(e, f)=(3, 3)$	-0.80048×10^3
$M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$	-0.11206×10^4

になる。

1) 小市-杭瀬下

小市に対しては正規分布的形状を、杭瀬下に対しては指数分布的形状を示す。

2) 小市-立ヶ花

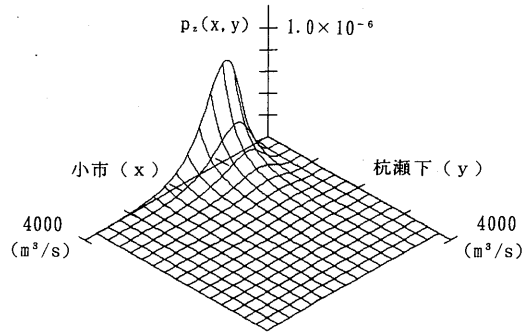
小市、立ヶ花双方に対して正規分布的形状を示す。

3) 杭瀬下-立ヶ花

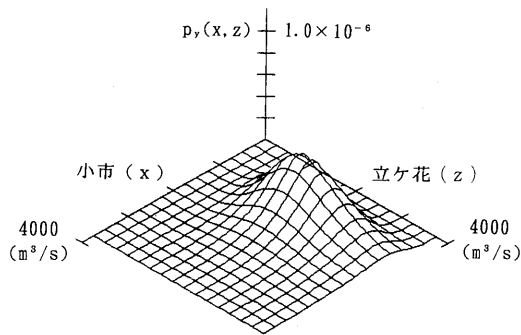
杭瀬下に対しては指数分布的形状を、立ヶ花に対しては正規分布的形状を示す。

また、図3は図2の分布と実データのメッシュ確率を示したものである。これより理論分布は実データに対して全体的な傾向は一応捉えているものと判断するが、データ数が少ないため厳密な考察は難しい。

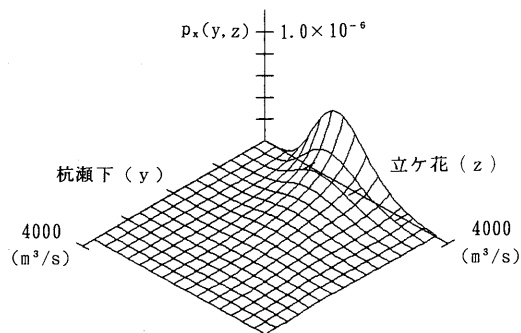
これに対し、図4、図5は従来の研究による統計モーメントのみを情報とした $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の密度関数立体図及びその分布と実データのメッシュ確率図である。図4



(a) 小市-杭瀬下



(b) 小市-立ヶ花



(c) 杭瀬下-立ヶ花

図2. $(e, f)=(1, 3)$ の分布の密度関数立体図.

においても上述の1)~3)の傾向は見られるが、分布形が緩やかな形状へと変形し、杭瀬下-立ヶ花において立ヶ花の値が大きいところでは杭瀬下方向への分布形状は正規分布的になってきている。また、図5においては図3のところでも述べたように、理論分布は実データに対して全体的な傾向は一応捉えているものと判断するが、データ数が少ないため、ここでも厳密な考察は難しい。

4.2 リターンピリオドの安定性

次に、L-Lで最も良い適合度を示した、 $(e, f)=(1, 3)$ の分布と、従来の統計モーメントのみ

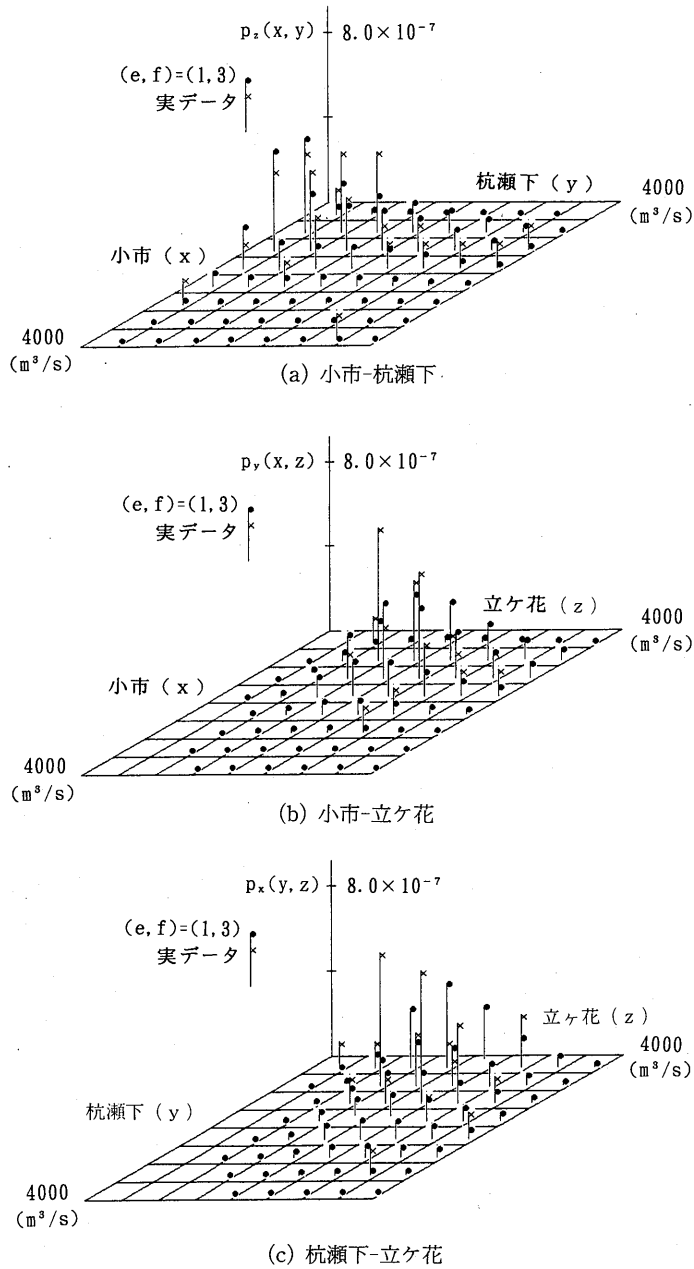
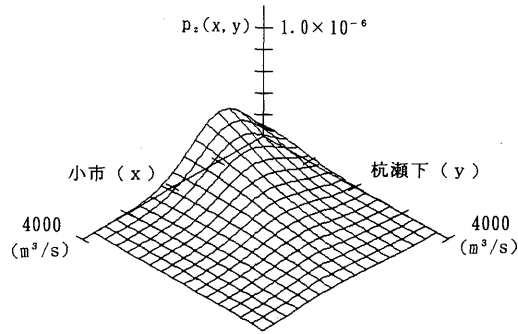
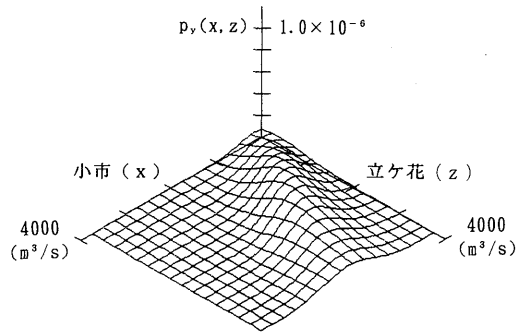


図3. $(e, f)=(1, 3)$ の分布とデータのメッシュ確率の比較.

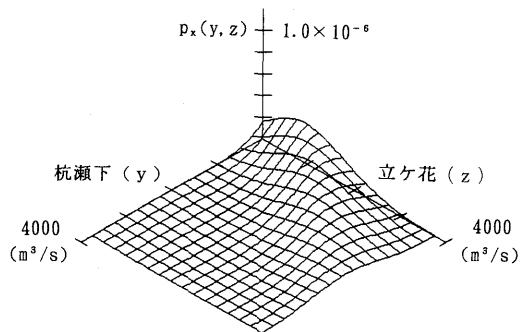
を情報とする $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布を用い、リターンピリオドの安定性について、比較検討を行った。具体的には、小市、杭瀬下、立ヶ花のある設定値を超過するリターンピリオドの安定性を見ることにより評価していくが、これは治水計画に重要な分布形の裾の部分の検討を行っていることに他ならない。設定値としては、現実的なリターンピリオドを用



(a) 小市-杭瀬下



(b) 小市-立ヶ花



(c) 杭瀬下-立ヶ花

図4. $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布の密度関数立体図.

いて比較したいため、小市、杭瀬下、立ヶ花それぞれの計画流量、4000, 5500, 9000 (m^3/s) の8割である、3200, 4400, 7200 (m^3/s) を採用することとした。表2は採用した分布の平均、標準分散、及び変動係数の計算結果である。結果を見ると、用いたデータセットに対しては、 $(e, f)=(1, 3)$ の分布の変動係数が0.150であるのに対し、 $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布では0.755と大きな値を示している。また、リターンピリオドの平均値を見ても、 $(e, f)=(1, 3)$ の分布では68年であるのに対し、 $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布では208年と大きな値となっている。すなわち、 $(e, f)=(1, 3)$ の分布では、流量が大きい方向に十分な尾を引き、分布の裾の部分の安定性についても、良い結果が得られていると判断できた。図6はリター

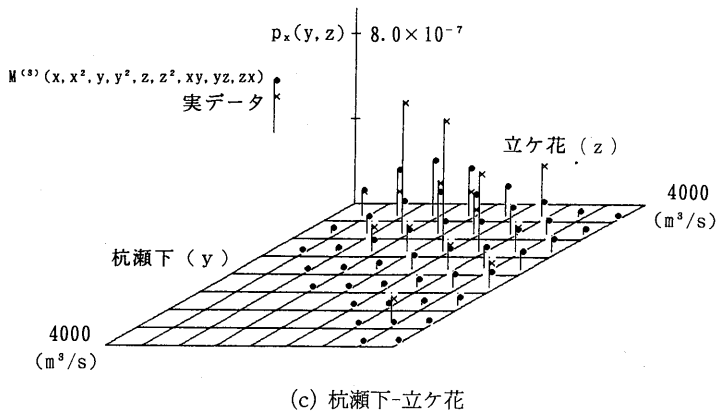
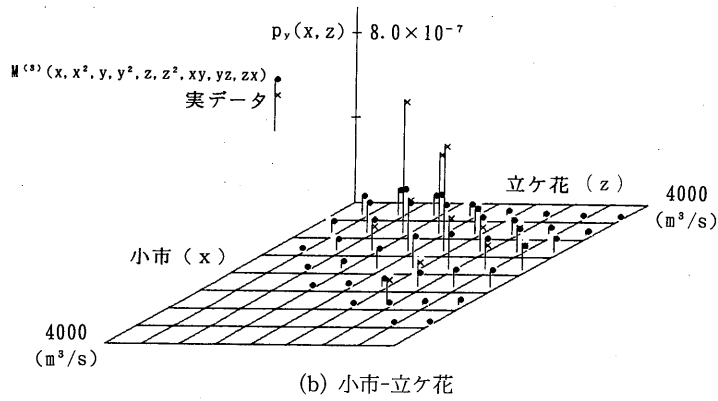
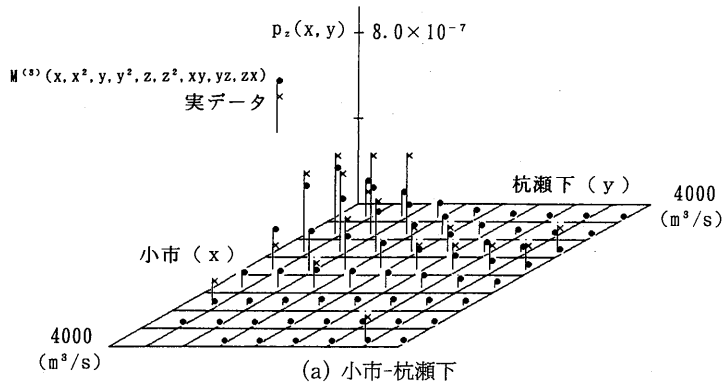


図5. $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布とデータのメッシュ確率の比較.

ンピリオドの変動を示したものである。抜いたデータ番号は年代の古い順に付けている。(e, f) = (1, 3) の分布の方が、1959年(データ番号6)と1983年データ(データ番号30)を取り除いたりターンピリオドの増加がかなり低く抑えられていることが分かる。

表 2. $(e, f)=(1, 3)$ の分布と $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布のリターンピリオドの平均, 標本分散及び変動係数.

分 布	平均(年)	標本分散	変動係数
$(e, f)=(1, 3)$	68	105	0.150
$M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$	208	24,562	0.755

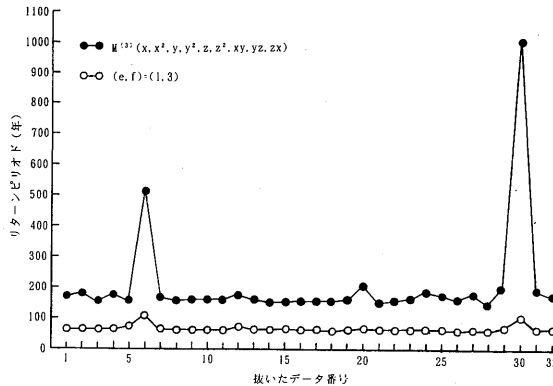


図 6. $(e, f)=(1, 3)$ の分布と $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布のリターンピリオドの安定性の比較.

5. ま と め

本論文では、多変数（3変数以上）頻度分析の分野で有効となる任意関数を情報とした多変数最大エントロピー分布のパラメータ同定法を示した。また、その中で3変数最大エントロピー分布への優れた情報の与え方を実データを用いて検討した。得られた成果を要約すると以下のようになる。

- 1) Wragg and Dowson (1970) の方法を拡張し、線形化した連立方程式を繰り返し計算することによって、任意関数を情報とした多変数最大エントロピー分布のパラメータが同定できた。
- 2) 分布形のデータに対する全体的な適合度の良否を L-L、密度関数立体図及びメッシュ確率図を用いて評価できた。
- 3) 今回用いた実データのケースでは、情報の与え方として統計モーメントのみならず、指数関数の期待値も取り入れた方が良好な適合度が得られることが分かった。更にその場合、 $(e, f)=(1, 3)$ の分布が L-L で最良の値を示した。
- 4) $(e, f)=(1, 3)$ の分布と $M^{(3)}(x, x^2, y, y^2, z, z^2, xy, yz, zx)$ の分布を用い、治水計画上重要なリターンピリオドの安定性について評価した。その結果、 $(e, f)=(1, 3)$ の分布の方がリターンピリオドの値が小さく安定していると判断された。

上述の成果は、従来3変数以上の場合正規分布を利用せざるを得なかった水文頻度分析の分野に、任意関数の期待値を情報としてエントロピーを最大にする分布を導入し、その分布はデータに対する全体的な適合度及びリターンピリオドの安定性の両者において有効性を示したため、水文学的には興味深く、インパクトを与え、また治水・利水計画に役立つものと考えられ

る。しかしながら、第1章で述べているように、本論文の第2章及び第3章で記述している統計学的方法には不備な点があるかも知れない。従って、著者等は、このことについて専門家の方々の統計学の見地からの妥当性の検討を期待している。

なお、本研究で用いたデータは建設省北陸地方建設局千曲川工事事務所から入手したものである。関心のある方はそちらに御請求頂きたい。

最後に、今後の課題としては次の事項が挙げられる。

- 1) 上述の統計学的に御検討頂いた内容に答える形で本論文を修正する。
- 2) 他の実データに対し、3変数最大エントロピー分布を適用し、最適な情報の与え方に法則性を見出す。
- 3) 簡単なパラメータ同定方法を開発し、パラメータ同定のための計算時間を短縮して、実用性を高める。

謝 辞

本論文に対して、文部省統計数理研究所の駒澤 勉先生、石黒真木夫先生、並びに「統計数理」編集委員会編集委員の先生方には有益な御助言と御指導を頂きました。また、本論文の査読に際して、査読者の方々から多くの貴重なコメントを頂きました。本論文の数値計算では、北野建設(株)の丸山貴彦君、並びに清水建設(株)の中川克洋君に御協力頂きました。最後になりましたが、明記して深甚なる謝意を表します。

参 考 文 献

- 石原安雄, 長尾正志 (1963). 計画高水流量の二変数確率論的研究, 京都大学防災研究所年報, **6**, 181-188.
- Krishnamoorthy, A.S. and Parthasarathy, M. (1951). A multivariate gamma-type distribution, *Ann. Math. Statist.*, **22**, 549-557.
- 長尾正志 (1975). 水文統計における多変数確率分布理論——二変数統計を中心として——, 水工学に関する夏期研修会講義集 (A コース), A-4-1-A-4-19, 土木学会水理委員会.
- 長沢敏夫 (1960). 複合確率と河川工事計画における応用について, 建設省直轄工事技術報告, **14**, 791-796.
- 寒川典昭, 荒木正夫, 佐藤健次 (1986). 多変数最大エントロピー分布とその基礎特性に関する研究, 土木学会論文集, **375/II-6**, 89-98.
- Sogawa, N., Araki, M. and Shimizu, K. (1989). Information suitable for annual flood and 2-variate maximum entropy distribution, *Pacific International Seminar on Water Resources System*, Tomamu, 103-116.
- 寒川典昭, 荒木正夫, 清水克彦 (1991). 月降水量に適した情報の与え方と2変数最大エントロピー分布, 信州大学工学部紀要, **69**, 25-39.
- Wragg, A. and Dowson, D. C. (1970). Fitting continuous probability density function over $[0, \infty)$ using information theory ideas, *IEEE Trans. Inform. Theory*, **IT-16**, 226-230.

Multivariate Maximum Entropy Distribution Using Information of
Expectations of Arbitrary Functions and Its Application to Frequency
Analysis at the Confluence of the Chikuma River

Noriaki Sogawa

(Faculty of Engineering, Shinshu University)

Masao Araki

(Kumamoto Institute of Technology)

Naoki Fukuzawa

(Fukuzawa Corporation)

This paper presents a parameter estimation method for a multivariate maximum entropy distribution (MMED) which has the information of expectation values of some arbitrary functions and this distribution is then applied to the frequency analysis of the confluence discharge of the Chikuma River (Nagano, Japan).

First, Wragg and Dowson's method is expanded to estimate the parameters of the MMED. Next, a log-likelihood (L-L) and the figure of probability in a mesh are used to judge the goodness-of-fit of the MMED to data, and a solid figure is drawn to grasp the whole shape of the MMED. Also, the stability of the tail of the MMED is evaluated by calculating the stability of the return period. Finally, the above methods are used to analyze the annual maximum discharge data at a gauging station after a confluence and the corresponding discharge data at two gauging stations before the confluence in the Chikuma River. The results obtained are as follows.

1) Distributions having the information of the expectations of exponential functions and statistical moments (distribution group A) resulted in high rankings while a distribution having only the information of statistical moments (distribution B) generated the worst ranking in the L-L calculations.

2) It was impossible to judge whether distribution group A or distribution B had a better fit to data because of the scarcity of data in the calculations of the figure of probability in the mesh.

3) Distribution group A had higher peaks and decreased more rapidly than distribution B in the solid figures.

4) The return periods of distribution group A were smaller and more stable than that of distribution B.

The results of applying the above method show that distribution group A produces better L-L and return period stability than distribution B.