

情報関数が最大エントロピー分布に 及ぼす影響の解析

寒川 典昭*・荒木 正夫**・大原 基秀***

Analysis of Effect of Information Function on Maximum Entropy Distribution

Noriaki SOGAWA*, Masao ARAKI**, and Motohide OHARA***

Synopsis

The effects of 4 information functions on a 1-variable maximum entropy distribution (1MED) were analyzed using simulation data from 6 assumed populations and annual maximum daily rainfall data. For this study, 6 gamma distributions with different parameters were taken as the populations and x^r , $1/x^r$, $(\ln x)^r$ and $\exp(-rx/M)$ ($M = E[X]$) functions were adopted as the four information functions. The calculation results using simulation data from the 6 gamma distributions could not fit the 1MED (obtained from information functions with $r=1,2$) to the population satisfactorily. The combined information of x^m and $\exp(-nx/M)$, or (m, n) -information, improved the fitness of the 1MED to the population in both the shape of the p. d. f. and the return period of the right tail. Furthermore, it was found that the (1,2), (1,3) and (1,4)-informations are particularly suitable for steep shaped populations having peaks in the left side, while the (1,2), (1,3), (2,3) and (2,4)-informations are best for flat form populations exhibiting peak shifts to the lower right. These 5 types of 1MEDs gave the most stable and realistic return periods for annual maximum daily rainfalls measured in the Nagano prefecture.

1. はじめに

水流量は一般に小標本であり、それが従う母集団分布の推定に多くの不確定さが存在している。高棹・池淵¹⁾は、水文系列の安定する統計的特性量を指定して最大エントロピー原理 (Principle of Maximum Entropy) により、水文系列が従う分布の特性を説明した。

著者ら²⁾は、高棹らの立場を水流量の頻度分析へと発展させ、Sonuga³⁾の2次モーメントまでを情報とした1変数最大エントロピー分布を任意次数のモーメントまで情報とした分布へと拡張した。さらに、著者らは相関のある多地点の水流量の同時頻度分析の重要性

* 社会開発工学科 助教授

** 熊本工業大学土木工学科 教授

*** 水資源開発公団

に鑑み、上述の分布を2変数⁴⁾、3変数以上の多変数⁵⁾、及び条件付き⁶⁾へと発展させるとともに、最大エントロピー分布の持つ諸特性を検討した⁷⁾。しかし、これら一連の研究では、適当な積分可能性をもつ関数を $g(s)$ (これを情報関数と呼ぶことにする)、確率変数を X とするとき、密度関数を確率変数 $g(X)$ の期待値で記述しているが、母集団の推定及び実データへの適用にあたっては、 $g(\cdot)$ として $g_r(x) = x^r$ 、すなわち統計モーメントによる情報に固定してきている。

最近著者ら⁸⁾は、この固定をはずし、情報関数と水工計画に用いられる実用的な指標である確率水文量あるいはリターンピリオドとの関連を把握することを試みはじめた。本研究はこの試行の一環として、個々の情報関係の挙動が密度関数に及ぼす影響を分析し、情報関数を組み合わせることの基本的な理由について検討して、母集団に対する全体的及び右裾のリターンピリオドの適合性がよく、しかも年最大日降水量のリターンピリオドが安定する情報関数を見出そうとしたものである。

2. 理論式の概説

第1章でも述べたように、確率変数を X 、適当な積分可能性をもつ任意の1変数関数を $g(\cdot)$ とする。いま情報関数 $g(\cdot)$ として $g_r(\cdot)$ ($r=1, 2, \dots, N$) を採用すると、1変数最大エントロピー分布 (1-Variable Maximum Entropy Distribution (1MED)) $p(\cdot)$ は次式のように推定される⁸⁾。

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

ここに、 λ_r ($r=0, 1, 2, \dots, N$) はラグランジュ乗数であり、推定分布 $p(\cdot)$ からみればその分布のパラメータである。(1)式は一般的な1MEDであり、 $g_r(\cdot)$ に上述した条件を満足する適当な関数を与えることにより、具体的な分布形が求められることになる。

(1)式のパラメータは次の $(N+1)$ 個の非線形連立方程式を解くことにより決定される⁸⁾。なお、本章では「 f_R 」を「 f 」と表わす。

$$\lambda_0 = \ln \left[\int \exp\left\{-\sum_{r=1}^N \lambda_r g_r(x)\right\} dx \right], \quad (2)$$

$$\int g_r(x) \exp\left\{-\sum_{r=1}^N \lambda_r g_r(x)\right\} dx = \nu_r \int \exp\left\{-\sum_{r=1}^N \lambda_r g_r(x)\right\} dx, \quad r=1, 2, \dots, N. \quad (3)$$

ただし、 ν_r は確率変数 $g_r(X)$ の期待値である。

ここで、(3)式の非線形方程式を λ_r ($r=1, 2, \dots, N$) に関して解くために、収束過程が明示でき、その収束度合をみながら補正量(残差)を修正することが可能であり、また以下の計算で十分な精度の解を得ることができた Wragg・Dowson⁹⁾の方法を採用する。

まず、 λ_r の近似値として α_r を与え、残差を ε_r とおくと、次式が成立する。

$$\lambda_r = \alpha_r + \varepsilon_r, \quad r=1, 2, \dots, N. \quad (4)$$

そこで、上式を(3)式に代入して、 $\exp\{-\varepsilon_r g_r(x)\}$ をテーラー展開し、 $(\varepsilon_r)^2$ 以上の微小量を無視して整理すると

$$\sum_{j=1}^N (A_{i,j} - \nu_i A_{0,j}) \epsilon_j = A_{i,0} - \nu_i A_{0,0}, \quad i=1,2,\dots,N \quad (5)$$

が得られる。ここに

$$A_{i,j} = \int g_i(x) g_j(x) \exp\left\{-\sum_{r=1}^N a_r g_r(x)\right\} dx, \quad i=0,1,2,\dots, \quad j=0,1,2,\dots \quad (6)$$

であり、 $g_0(x) = 1$ とする。

(5)式は残差 $\{\epsilon_1, \epsilon_2, \dots, \epsilon_N\}$ に関する N 元連立 1 次方程式であり、変数の数と方程式の数が一致するので解き得る。数値計算では、初期値 $\{a_1, a_2, \dots, a_N\}$ を与え、残差について(5)式を解く計算を残差が十分小さく (10^{-8} 以下) なるまで繰り返すことにより、解 $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ に実用的にはほぼ等しい近似値を得ることができる。 λ_0 は求めた λ_r ($r=1, 2, \dots, N$) を(2)式に代入して算定される。

3. 既存の母集団に対する適合性

3.1 同種の関数を個々に情報関数とした場合

本節では、 $g_r(\cdot)$ として、統計モーメントに対応する x^r の他に、 $1/x^r$ 、 $\exp(-rx/M)$ ($M: x$ の平均)、 $(\ln x)^r$ の代表的な 4 つの関数を個々に情報関数とする場合について検討する。なお、ここではパラメータ数をおさえるために、 $r=1,2$ を採用する。

まず、母集団として、次式で表現されるガンマ分布を仮定した。

$$p(x) = \frac{1}{\beta^{\alpha+1} \Gamma(\alpha+1)} x^{\alpha} \exp(-x/\beta) \equiv G(\alpha, \beta), \quad \alpha > 0, \beta > 0, x > 0. \quad (7)$$

今回は、指数分布 $G(0, 1)$ 、及び左にピークを持つ分布 $G(1, 0.7)$ からしだいに右へピークを移動し、フラットになった $G(1, 1)$ 、 $G(3, 1)$ 、 $G(3, 2)$ 、 $G(3, 3)$ の 6 つを用い、これより 1000 個の乱数を発生させて ν_r を計算し、その情報に基づいて 1MED を求めた。データ数を 1000 個にしたのは、まず適切な関数を選定するため、ほぼ母集団に近い情報が得られる状態を設定したためである。

図 1 は母集団分布とヒストグラムに対する推定分布の全体的な形状の適合性をみた 1 例である。母集団 $G(0, 1)$ は指数分布であり、単調減少している。この場合、 x^r で情報を与えた 1MED は、すべての領域にわたってほとんど母集団に合致しているが、他の関数では十分に母集団を再現することができなかった。

$G(0, 1)$ 以外の母集団分布に対する適合では、以下に述べる同様な傾向を示した。

x^r の場合は、母集団分布及び他の関数で情報を与えた推定分布に比べてフラットになり、右側に長く尾を引かない。逆に、 $1/x^r$ の場合は、かなり全体的に左側に移動して母集団から離れたとがった分布となり、右側に長く尾を引く。他の 2 つの場合は、上述の 2 つの中間的な分布となり、 $1/x^r$ 、 $\exp(-rx/M)$ 、 $(\ln x)^r$ 、 x^r の順にピークが下がりフラットな形状を示す。

そこで次に、全体的に比較的良好な適合性を示した、 x^r 、 $\exp(-rx/M)$ 、 $(\ln x)^r$ の 3 つの場合について、次章の極値水文量への適用において重要となる右裾の適合性をみるために、指定した x の値に対して求めたリターンピリオドの 1 例を図 2 に示した。

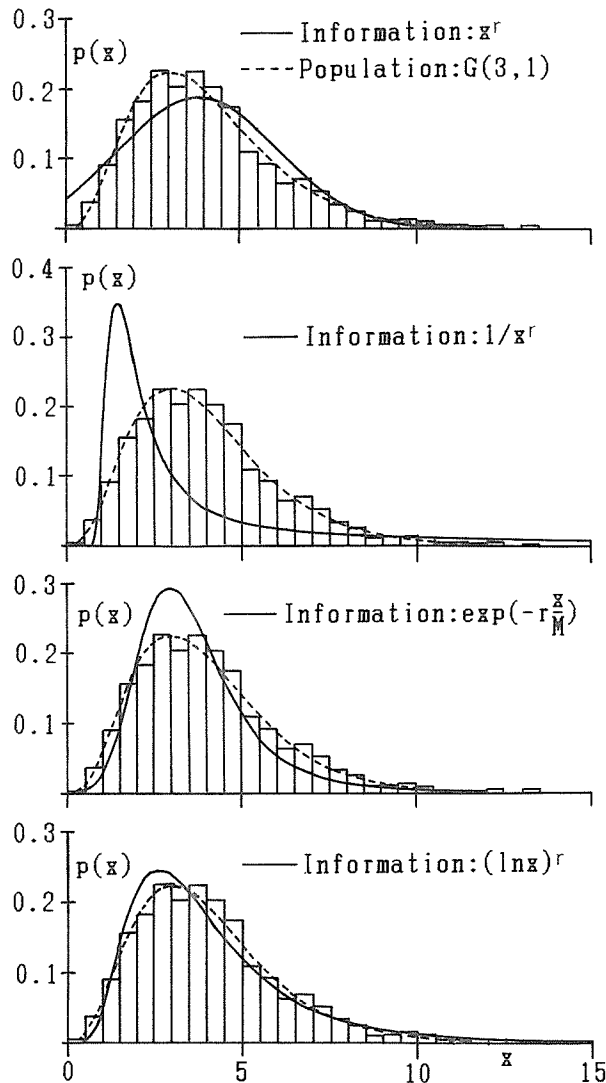


Fig. 1 Fitting to the histogram and the p. d. f. curve in $G(3, 1)$ (Information from the same kind of functions, the number of data: 1000).

母集団を $G(0, 1)$ とした場合、やはり x^r の場合の適合性が他に比べてよい。また、他の母集団については、母集団の種類にかかわらず、 x の大きな値のところでのリターンピリオドは、 x^r の場合は母集団より大きく、他の 2 つの場合は小さな値となった。

3.2 異種関数を組み合わせて情報関数とした場合

前節での結果から、 x^r と、 $(\ln x)^r$ か $\exp(-rx/M)$ のどちらか一方を組み合わせた分布は、単独の関数から得られた分布より全体的な適合性、及びリターンピリオドの適合性

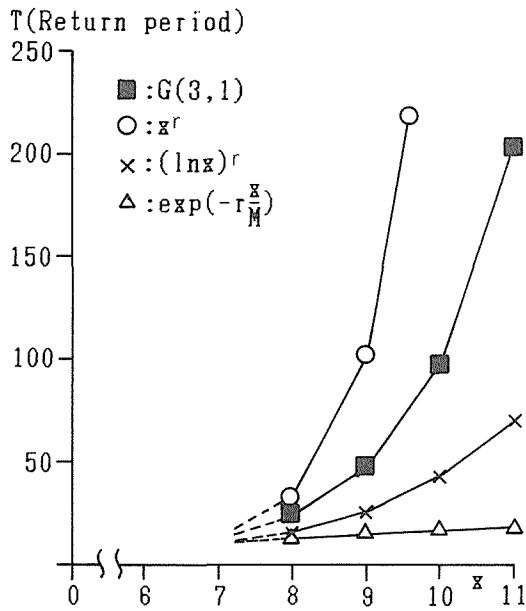


Fig. 2 Fitting to the return periods of right tail in $G(3, 1)$ (Information from the same kind of functions, the number of data : 1000).

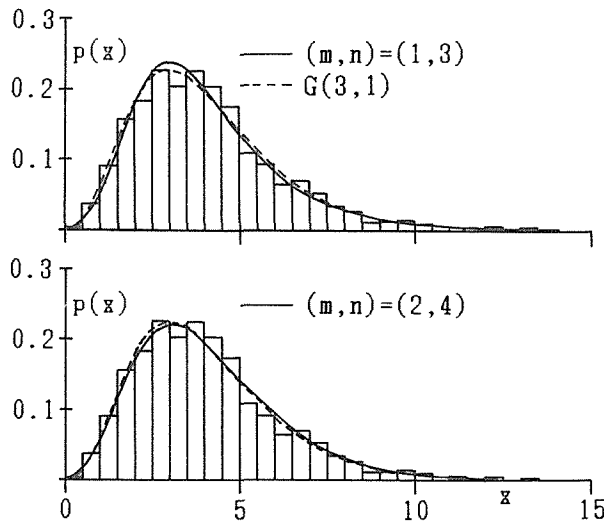


Fig. 3 Fitting to the histogram and the p. d. f. curve in $G(3, 1)$ (Information from the different kind of combined functions, the number of data : 1000).

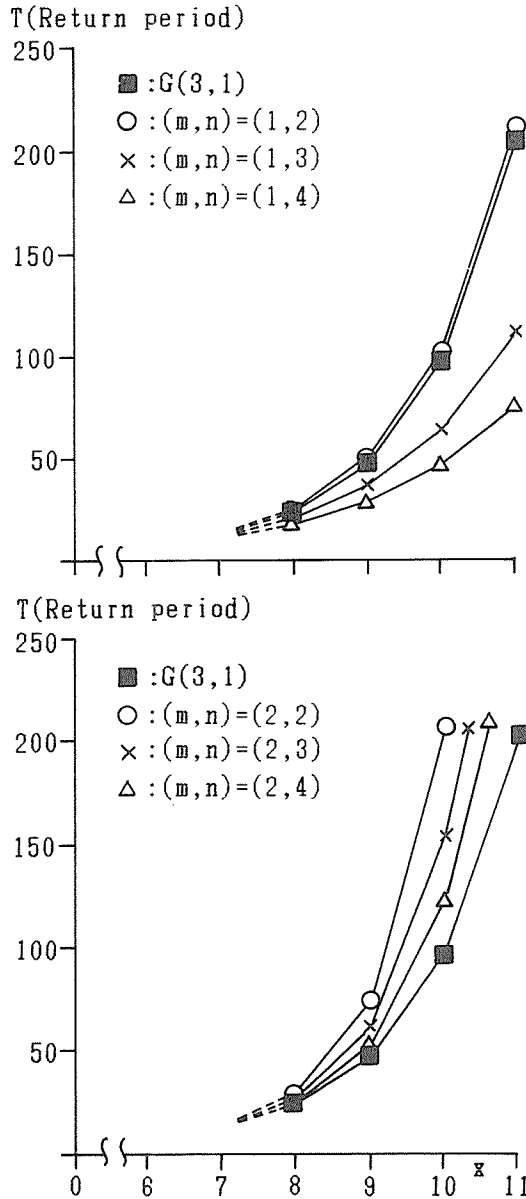


Fig. 4 Fitting to the return periods of right tail in $G(3, 1)$ (Information from the different kind of functions, the number of data: 1000).

の点で優れていることが予想される。ところが、 $y=(\ln x)^r$ は r が奇数の時負の領域を含み、 r が偶数の時固定した y の値に対して2つの x の値が存在するため、他の関数と挙動を異にする。従って本節では、 x^r と $\exp(-rx/M)$ を組み合わせる情報関数として用いる方法について検討する。

さてこの場合、(1)式より、1MEDは次式のように表わされる。

$$p(x) = \exp\{-\lambda_0 - \lambda_1 x^m - \lambda_2 \exp(-nx/M)\}. \quad (8)$$

ただし、 m 、 n は任意の正の整数である。

ここでは、この分布の特性を知るために、上式中の (m, n) の値を変化させながら $G(1, 1)$ と $G(3, 1)$ を母集団とした場合について前節と同様な検討を行った。ただし、 $n=1$ とした場合は、パラメータ同定がなされなかったため、用いる分布を $(m, n) = (1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4)$ の12ケースとした。

図3は、母集団分布とヒストグラムに対する推定分布の全体的な形状の適合性を、図4はリターンピリオドの適合性をみた1例である。一般に、分布形は m の値が大きくなると、ピークの位置が少しずつ右下に下がり、フラットになる。また、あまり右側に尾を引かなくなるため、リターンピリオドが大きくなる。逆に、 n の値が大きくなると、ピークの位置が少しずつ左上に上がり、とがって全体的に左側に移動してくるが、右側に長く尾を引くため、リターンピリオドが小さくなる。しかし、 m の値が3、4となると、母集団分布やヒストグラムに対する適合性が悪くなり、 x の大きな値に対するリターンピリオドも常に真値より大きくなる。

以上の考察より、(8)式の分布の特性が明らかになったが、母集団分布に対するリターンピリオドの適合性から判断して、 $G(1, 1)$ のような左側にピークを持ち、とがった母集団に対しては、 $(m, n) = (1, 2), (1, 3), (1, 4)$ を、また $G(3, 1)$ のようなフラットな母集団に対しては $(m, n) = (1, 2), (1, 3), (2, 3), (2, 4)$ を用いるとよい。

3.3 同数の異種データによる母集団推定

ここでは、母集団を $G(1, 1)$ と $G(3, 1)$ とし、これより50個と100個の乱数を10組ずつ発生させた。また、採用分布として、前節で適合性のよかった $(m, n) = (1, 2), (1, 3), (1, 4), (2, 3), (2, 4)$ を用い、これらの分布の、ヒストグラムと母集団分布に対する全体的な適合性とリターンピリオドの適合性を検討した。

以下では、明確な特徴が表われた次の4つの場合に分類して適合性を考察する。

1) 平均、分散が共にほぼ母集団に等しい場合

このとき、推定分布は、ヒストグラムに対してはもとより母集団分布に対してもよい適合性を示す。リターンピリオドについては、 $G(1, 1)$ に対して $(m, n) = (1, 3), (1, 4)$ 、 $G(3, 1)$ に対して $(m, n) = (1, 2), (1, 3), (2, 3)$ の場合が、よく適合している。

2) 分散が母集団とほぼ等しく、平均が母集団より大きい場合

このとき分布形は全体的に母集団より右にずれる。このため、1) に比べてリターンピリオドの値が全体的に小さくなり、 $G(1, 1)$ に対しては $(m, n) = (1, 2), (2, 4)$ 、 $G(3, 1)$ に対しては $(m, n) = (1, 2), (2, 3), (2, 4)$ がよい適合性を示すようになる。

3) 平均が母集団とほぼ等しく、分散が母集団より大きい場合

このとき、 $(m, n) = (1, 3), (1, 4)$ は母集団よりとがり、 $(m, n) = (1, 2), (2, 3), (2, 4)$ は、母集団よりフラットになる。リターンピリオドの値はこの場合も全体的に小さくなり、 $G(1, 1)$ 、 $G(3, 1)$ のどちらの母集団に対しても、 $(m, n) = (1, 2), (2, 3),$

(2, 4) がよく適合している。

4) 平均が母集団とほぼ等しく、分散が母集団より小さい場合

このとき、どの推定分布もとがった形状を示し、長く尾を引かない。そのため、リターンピリオドの値が全体的に大きくなり、比較的長く尾を引く性質をもつ $(m, n) = (1, 3), (1, 4)$ が、 $G(1, 1), G(3, 1)$ のどちらの母集団に対してもよく適合する。

4. 実データへの適用

4.1 ヒストグラムに対する適合性

ここでは、3.2でよい適合性を示した(8)式を長野、上田、松本、北牧、飯山（いずれも長野県内）の年最大日降水量に適用した。なお、 (m, n) の値は、前章の考察より、 $(1, 2), (1, 3), (1, 4), (2, 3), (2, 4)$ の5ケースを採用するとともに、比較のため $x^r (r=1, 2)$ を情報関数とする方法での適用も行った。

図5は、全データ数に対するヒストグラムと推定分布の1例である。この図を見ると、 x^r の場合の推定分布は、(8)式の分布に比べてフラットで右側にピークを持ち、長く尾を

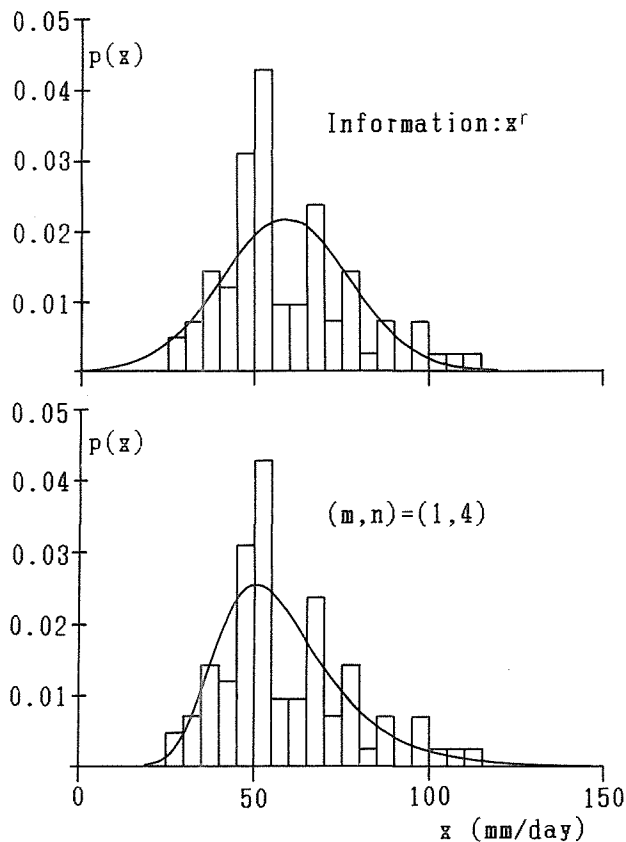


Fig. 5 Fitting to the histogram in annual maximum daily rainfall at Nagano.

引かない。また、(8)式の分布の場合は、3.2でも記述したように、 m が大きくなるとピークが右下に下がってフラットになり、 n が大きくなると全体的に左側に移動し、ピークが左上に上がってとがった分布形になっている。

4.2 リターンピリオドの評価

図6は、データを年代の古い順に並べ、前半分のデータの最大値のリターンピリオドが、前半分のデータに1つずつデータを追加することにより、どのような変動を示すかを図化した1例である。この図からもわかるように、 x^r で情報を与えた場合は、どの地点においても、リターンピリオドの変動幅、値ともに大きい。それと比較して(8)式の場合はこれらの値が小さく、特に m が小さいほど、また n が大きいほど、その特徴が顕著に表われている。このことは、(8)式の分布は右側に長く尾を引き分布形状が比較的固定されているためと推察される。

一方、表1は、全データ中の最大値のリターンピリオドを全データを用いて推定された1MEDより求めたものである。 x^r の場合は、データ数、最大値、平均値から判断して、最大値のリターンピリオドとしては大きすぎる。(8)式の場合は、 m が小さく n が大きい程、リターンピリオドの値は小さくなっているが、飯山の場合を除いてはどの場合も現実的な値を示している。飯山では、全てのリターンピリオドが非常に大きな値をとっている。

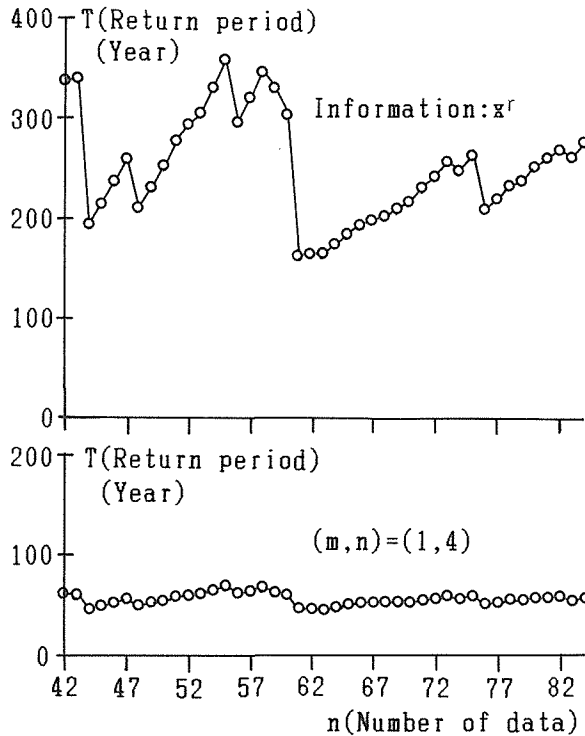


Fig. 6 Stability of return periods for maximum value 108.2 (mm/day) (Annual maximum daily rainfall at Nagano).

Table 1 Return periods for maximum values.

Rain gauge station	Number of data	Maximum value (mm/day)	Mean value (mm/day)	x^r	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)
Nagano	84	113.9	58.5	730	183	118	85	216	163
Ueda	79	119.0	62.4	183	79	58	46	87	73
Matumoto	72	118.6	67.4	128	62	48	39	66	56
Kitamaki	54	134.0	75.2	86	39	30	25	43	36
Iiyama	77	156.5	63.5	—	3430	1289	639	7181	4144

これは、データの平均値63.5(mm/day) に対して、最大値156.5 (mm/day) が、きわだって大きいためである。

なお、 x^r で情報を与えた前章と本章の結果は、さらに高次モーメントまで情報として利用することにより改善が期待されるが、ここでは他の分布とパラメータ数をそろえるため2次モーメントまでしか利用しなかったことを付記しておく。

5. 結 論

本章では、パラメータの数を2個に固定して、母集団の分布形とヒストグラムに対する全体的な形状の適合性がよく、かつ分布形の右裾の部分のリターンピリオドの適合性にすぐれ、さらにリターンピリオドが安定した推定分布を得るために、情報関数が1MEDに与える影響を分析した。以下に得られた成果を要約する。

1) 6種類のガンマ分布を母集団とし、情報関数として x^r , $1/x^r$, $\exp(-rx/M)$ ($M=E[x]$), $(\ln x)^r$ ($r=1, 2$) を取り上げて母集団推定を行うと、上述の6つの母集団の1つである指数関数母集団に対しては、 x^r の場合が全体的な形状でもリターンピリオドの点でもよい適合性を示すが、その他の母集団に対しては、特にリターンピリオドの点でどの関数の場合も十分な精度の推定値を得ることができなかった。

2) x^m と $\exp(-nx/M)$ の2つの関数を組み合わせて情報を与えると、 $G(1, 1)$ のような左側にピークを持ち、とがった母集団に対しては、 $(m, n)=(1, 2), (1, 3), (1, 4)$ を、また $G(3, 1)$ のようなフラットな母集団に対しては $(m, n)=(1, 2), (1, 3), (2, 3), (2, 4)$ を用いると母集団の分布形に対する全体的な適合性のみならず、リターンピリオドの点でもよい適合性が得られた。

3) $G(1, 1)$ と $G(3, 1)$ の母集団から、50個、100個の乱数を10組ずつ発生させ、 $(m, n)=(1, 2), (1, 3), (1, 4), (2, 3), (2, 4)$ で母集団推定を行うと、データの平均と分散の真値に対する大小関係から、適合する (m, n) の値を少数個に絞ることができた。

4) 長野県内5地点で得られた年最大日降水量データを用いて、 $(m, n)=(1, 2), (1, 3), (1, 4), (2, 3), (2, 4)$ で母集団推定を行うと、どの推定分布も比較的安定して現実的なリターンピリオドを与えることが明らかになった。

なお、ここで取り上げた1MEDは、分布形を固定しないかわりに情報関数の選択に主観が入り、それに分布形が依存して決定されるという点で、両者の間にトレードオフ的な

意味あいが存在するが、適切な情報関数を用いることにより、この分布は既存の分布よりすぐれた効果を発揮することが予想され、対象とする水文量（例えば年最大日降水量といった特定の時間単位の水文量）ごとに、情報関数を少数個に絞ることができるという点からも、1MED はかなり有効な分布であると言えよう。しかも、上述の利点を保持したまま容易に多変数に拡張され得るということは、従来の分布にない特徴である。

参 考 文 献

- 1) 高棹琢馬, 池淵周一: エントロピー的にみた降雨・流出変換特性とそのモデル化, 京都大学防災研究所年報, 第23号, B-2, pp. 193~209 (1980年).
- 2) 寒川典昭, 荒木正夫: 水文事象の頻度分析への MEP 導入について, 土木学会論文報告集, 第335号, pp. 89~95 (1983年).
- 3) Sonuga, J. O.: Principle of maximum Entropy in hydrologic frequency analysis, Jr. of Hydrology, Vol. 17, pp. 177~191 (1972).
- 4) 寒川典昭, 荒木正夫, 寺島 彰: 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, pp. 397~402 (1984年).
- 5) 寒川典昭, 荒木正夫, 佐藤健次: 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, 第375号/II-6, pp. 89~98 (1986年).
- 6) Sogawa, N., M. Araki and T. Imai: Studies on multivariate conditional maximum entropy distribution and its characteristics, JHHE, Vol. 4, No. 1, pp. 79~97 (1986).
- 7) 寒川典昭, 荒木正夫, 大原基秀, 森川 智: 最大エントロピー分布の諸特性の検討, 第31回水理講演会論文集, pp. 185~190 (1987年).
- 8) 寒川典昭, 荒木正夫, 船橋太道: 年最大降雨に適した情報と母数の数と最大エントロピー分布, 土木計画学研究・論文集, 第7号, pp. 91~98 (1989年).
- 9) Wragg, A. and D. C. Dowson: Fitting conditional probability density function over $(0, \infty)$ using information theory idea, IEEE Transaction on Information Theory, IT-16, pp. 226~230 (1970).