II―11　　Shannon流情報理論による確率密度関数の決定

信州大学工学部　正員　荒木　正夫
同　上　正員　○寒川　典昭
同　上　学生員　小山　金利

§1. はじめに

SONUGA[1]の水文事象の頻度分析に関する理論はエントロピー最大化法を導入した理論の1つであるが，解法が煩雑でしかもラグランジュ乗数に関して3変数族以上に拡張することが難しい。本研究は情報理論の分野でWRAGGら[2]によって展開された多変数族に関する確率密度関数のfitting理論を導入し，その特性を検討しながら水文事象への適用をはかろうとするものである。

§2. 理論式の算定

連続変数xについて$[0,\infty)$でのエントロピーHと原点のまわりのr次のモーメントμ'_rは次式で示される。

$$H=-\int_0^\infty p(x)\,dx \qquad \ldots(1)$$

$$\mu'_r=\int_0^\infty x^r p(x)\,dx \quad r=0,1,\ldots,N \qquad \ldots(2)$$

ただし，$p(x)$は密度関数であり，その性質から$p(x)\geq 0$，$\mu'_0=1$である。さて，(2)式を制約条件として(1)式を最大にする$p(x)$を求めることは(2)式でうける制約内でできるだけ一様となる$p(x)$を求めることである。換言すれば，xが(2)式で与えられる情報以外はできるだけランダムに発生するように$p(x)$を決定することである。我々は水文事象の密度関数の決定にこのような立場をとることが妥当であり，自然であろうと考える。そこで上記の問題をラグランジュの未定乗数法で解くと$p(x)$の推定値$\hat{p}(x)$は

$$\hat{p}(x)=A\exp\Bigl(-\sum_{r=1}^{N}\lambda_r x^r\Bigr) \qquad 0\leq x<\infty \qquad \ldots(3)$$

となり，上式に含まれるラグランジュの未定乗数λ_rは

$$\mu'_r\int_0^\infty \exp\Bigl(-\sum_{r=1}^{N}\lambda_r x^r\Bigr)dx=\int_0^\infty x^r\exp\Bigl(-\sum_{r=1}^{N}\lambda_r x^r\Bigr)dx \qquad r=1,2,\ldots,N \qquad \ldots(4)$$

なる非線形方程式を解くことによって求められる。ただし，(1)，(2)式で与えられる問題が(3)，(4)式の解を持つためには最大次数であるN次モーメントに対して次の制約条件を満たさなければならないことが知られている[3]。

$$\mu'_N\leq\mu'_{N,\max}=\int_0^\infty x^N\exp\Bigl(-\sum_{r=0}^{N-1}\lambda_r x^r\Bigr)dx \qquad \ldots(5)$$

$$\mu'_N=\mu'_{2i}>\mu'_{2i,\min}=-\begin{vmatrix}1 & \mu'_1 & \cdots & \mu'_i\\ \mu'_1 & \mu'_2 & \cdots & \\ \vdots & \vdots & & \\ \mu'_i & & & 0\end{vmatrix}\cdot\begin{vmatrix}1 & \mu'_1 & \cdots & \mu'_{i-1}\\ \mu'_1 & \mu'_2 & \cdots & \\ & & & \\ \mu'_{i-1} & & & \mu'_{2(i-1)}\end{vmatrix}^{-1} \qquad \ldots(6)$$

ここで，λ_1，λ_2の2変数族を考えてみよう。このとき(3)式は

$$\hat{p}(x)=A\exp\bigl\{-(\lambda_1 x+\lambda_2 x^2)\bigr\} \qquad 0\leq x<\infty \qquad \ldots(7)$$

となり，上式に含まれるAは次式で求められる。

$$A=(2\lambda_2)^{1/2}B(\xi) \qquad \cdots(8) \qquad\qquad B(\xi)=\frac{\phi(\xi)}{1-\Phi(\xi)} \qquad \cdots(9)$$

ここに，ϕ，Φは正規分布の密度関数，分布関数を示し，$\xi=\lambda_1/(2\lambda_2)^{1/2}$ である。

又，$p(x)$ の標準化は次のようにしてなされる。

$$x=(\mu_1'/\mu_0')y \qquad \cdots(10) \qquad\qquad p(x)=\{(\mu_0')^2/\mu_1'\}\,q(y) \qquad \cdots(11)$$

従って $q(y)$ は原点に関して r 次のモーメント

$$m_r=(\mu_0'/\mu_1')^r(\mu_r'/\mu_0') \qquad \cdots(12)$$

を持ち，$m_0=m_1=1$ となる。

§3. 適用結果と考察

ここでは2変数族の場合の適用結果を示す。Fig.1はm_2を従属変数として描いたλ_1，λ_2の変化特性である。$m_2<1.5$ の区間では両者とも大きな勾配をもち，密度関数がモーメントの微小変化に対して大きく変動することがうかがわれる。このようにm_2に対するλ_1，λ_2が得られると任意分布をfittingすることが可能となるが一例として密度関数が

$$p(x)=630x^4(1-x)^4 \qquad \cdots(13)$$

で示されるベータ分布のfittingを考察する。Fig.2がその適合結果でありこのような分布では2次モーメントまでで十分よい精度が得られている。

Fig.3はm_2をパラメータとしてyと確率年との関係を記述したものである。yが1と2の区間で各曲線が交わりyが1以下ではm_2が大きいほど2以上では小さいほど確率年が大きくなっている。Fig.4は前出のベータ分布を母集団として任意の個数の乱数を20回発生させて得たλ_2の平均値と標準偏差である。個数が60個付近から標準偏差が小さくなりλ_2の平均が真値に近づいてきていることは確率年を議論するときのデータ数に対する情報を与えるものである。

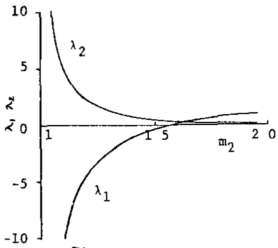

Fig. 1 m_2とλ_1，λ_2の関係

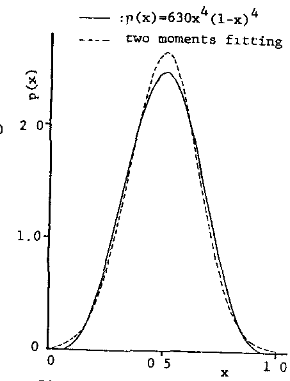

Fig. 2 2次モーメントまでを使用したベータ分布のfitting

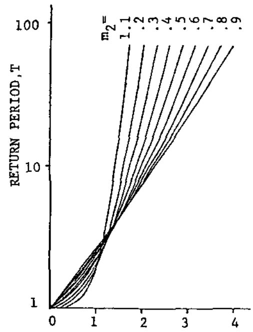

Fig. 3 m_2をパラメータとしたyと確率年との関係

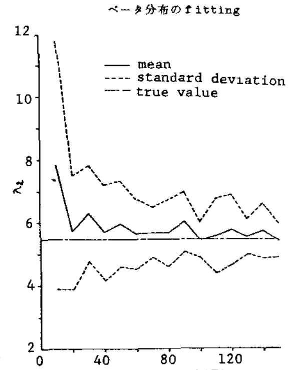

Fig. 4 λ_2の安定性

§4. あとがき

本稿では水文事象の確率密度関数の決定にエントロピー最大化法を採用し，2変数族を例にあげてベータ分布のfitting及び確率年を議論したが，多変数族への適用をはかることも可能である。これについては講演時に述べることにする。

1)SONUGA : Principle of Maximum Entropy in Hydrologic Frecency Analysis ; Journal of Hydrology 17 , 1972
2)WRAGG & DOWSON : Fitting Continuous Probability Density Function Over [0,∞) Using Information Theory Ideas ; IEEE TRANSFORMATION ON INFORMATION THEORY , MARCH 1970.
3)EINBO : On the Existence of a Class of Maximum-Entropy Probability Density Function ; IEEE TRANSFORMATION OF INFORMATION THEORY,NOVENBER 1977.