II－5

水文データのもつ情報量の評価

信州大学工学部　正　員　荒木正夫　　信州大学工学部　正　員　寒川典昭
信州大学大学院　学生員○船橋太道　　信州大学工学部　　　　清水克彦

1．はじめに

水文データのもつ情報量とは何か，それをいかに数量化するか，これが本稿の目的である。

一般に情報量とは，ある対象に提供された知識の量であり，換言すれば不確定さの減少量と言える。したがって，情報理論の分野で議論されているように

（情報量）＝（事前の不確定さ）－（事後の不確定さ）

とするなら，それが抽象的な情報量の定義になる。問題は不確定さの定式化であるが，Shannonのエントロピーはこの数量的表現として適している。

さて，確率分布による水文統計を問題にするとき，水文データのもつ情報量の対象は密度関数のもつ母数とみなすことができる。そこで，母数を確率変数とし，データを得る前後で母数のエントロピーを測り，両者の差を計算することからそのデータのもつ情報量が求まる。

尚，採用する水文データは，基礎確率変数が確率分布の中で最も基本となる正規分布に従う場合，すなわち，正規情報が利用可能な場合に限り，年降水量とした。

2．理論式[1]

いま，確率変数$\tilde{x}$があり，それは平均θ，分散v（精度h＝1／v）をもって正規分布をするものと仮定する。このとき$\tilde{x}$の実現値を知ることを正規情報を得るという。

a）平均未知，分散既知の場合

平均θが未知で分散vが既知の場合の$\tilde{\theta}$の事前確率密度関数$\xi(\theta)$は，文献1）の（1）式で与えられ，$\tilde{x}=x$（分散既知の正規情報）が得られた後の事後確率密度関数$\xi(\theta\mid x)$，及びそのパラメーターは同（2），（3）式において，n＝1として得られる。

そこで，データxのもつ情報量$I(\theta\mid x)$は次式となる。

$$I(\theta\mid x)=H(\xi(\theta))-H(\xi(\theta\mid x))$$
$$=\ln\left(\frac{n_{max}/2+1}{n_{max}/2}\right) \qquad (1)$$

ここで，H（・）はエントロピーを表す。

b）平均未知，分散未知の場合

平均θ，精度h＝1／vが共に未知の場合の（$\tilde{\theta}$，$\tilde{h}$）の同時事前確率密度関数$\xi(\theta, h)$は，文献1）の（4）式の正規－ガンマ分布で与えられ，そのパラメーターは同（18）式で与えられる。それ故，$\tilde{x}=x$（分散未知の正規情報）が得られた後の同時事後確率密度関数$\xi(\theta, h\mid x)$もやはり正規－ガンマ分布で与えられ，そのパラメーターは同（16），（17）式においてn＝1を代入することにより算定される。

そこで，データxのもつ情報量$I(\theta, h\mid x)$は次式となる。

$$I(\theta,h \mid x) = H(\xi(\theta,h)) - H(\xi(\theta,h \mid x))$$

$$= \ln\frac{A_2}{A_1} + B_1\left\{1-\Psi\left(B_1+\frac{1}{2}\right)+\ln C_1\right\} - B_2\left\{1-\Psi\left(B_2+\frac{1}{2}\right)+\ln C_2\right\} \qquad (2)$$

ここに，

$$\begin{cases} A_1 = \dfrac{n'}{(2\pi)^{1/2}}\cdot\dfrac{1}{\Gamma(\nu'/2)}\cdot\left(\dfrac{1}{2}\beta'\nu'\right)^{\frac{\nu'}{2}}, & A_2 = \dfrac{n''}{(2\pi)^{1/2}}\cdot\dfrac{1}{\Gamma(\nu''/2)}\cdot\left(\dfrac{1}{2}\beta''\nu''\right)^{\frac{\nu''}{2}} \\ B_1 = \dfrac{1}{2}\beta'\nu' & , B_2 = \dfrac{1}{2}\beta''\nu'' \\ C_1 = \dfrac{1}{2}(\nu'-1) & , C_2 = \dfrac{1}{2}(\nu''-1) \end{cases}$$

３．実測データへの適用と考察

a）平均未知，分散既知の場合

（１）式から明らかなように，データのもつ情報量はデータの値によらず一定である。つまり，どのデータも平均推定という立場では同量の情報をもっていることになる。Fig.１は長野県内の１９の観測所について，(１)式の計算結果を示したものであり，観測所ごとにデータのもつ情報量の値が読み取れる。

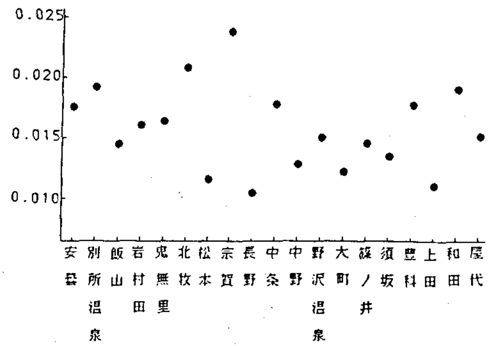

Fig.１　観測地点と情報量

b）平均未知，分散未知の場合

Fig.２は，N（０，1^2）に従う乱数を発生させ，個々の乱数が母数推定にどれほどの情報量をもつかを（２）式よりシミュレートしたものである。この図から，平均値より離れた乱数（データ）ほど大きな情報量をもっていることがわかる。Fig.３には，Fig.１の１９の観測所について同様の検討をしたものの１例（松本）を示した。Fig.２と同じ考察が得られる。以上より，平均・分散を同時に推定する立場からみると，平均から離れているデータほどその推定精度を高めることがわかる。

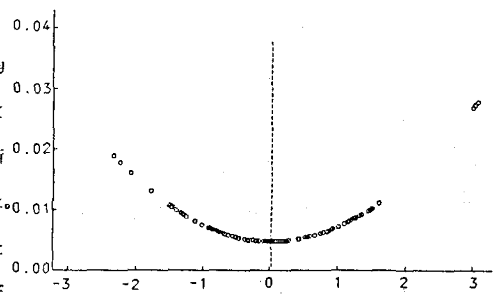

Fig.２　正規乱数と情報量

４．あとがき

本稿では，正規情報が得られる場合に，個々のデータが母数推定にどれほどの情報量をもつかを検討した。今後，情報量の値の意味を明確にしていくとともに，事前状態の与え方，及び正規以外の情報が得られる場合等の検討を進めたい。

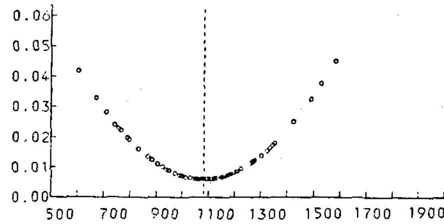

Fig.３　年降水量（松本）と情報量

１）寒川典昭・荒木正夫・渡辺輝彦：確率分布の母数推定の不確定性評価法，土木学会論文集，第３７５号／Ⅱ－６，pp133 ～ 141，1986