

水文データのもつ情報量の評価

寒川典昭* 荒木正夫** 船橋太道***
清水克彦****

(昭和63年3月29日受理)

Evaluation of Information Content of Hydrological Data

Noriaki SOGAWA,* Masao ARAKI,** Taido FUNAHASHI,***
and Katsuhiko SHIMIZU****

Theoretical equations were derived for the evaluation of information content of hydrological data, i. e., the degree of decrease in uncertainty of prior distribution, and applied to a simulation data and an annual rainfall data. The theoretical equations refer to a specific case where the basic random variate, expressive of hydrological data, obeys a normal distribution $N(\mu, \sigma^2)$ and to two conditions A (μ : variate, σ^2 : known) and B (μ, σ^2 : variate). The derivation made use of the concept of Shannon's entropy. The simulation data was based on $N(0, 1^2)$ and the annual rainfall data was obtained by 19 rain-gauge stations in the Nagano prefecture. According to the application of the rainfall and simulation data, for condition A the information content depends only on the number of data used for the formation of prior distribution, whereas for condition B it depends not only on the characteristics of the data used for the formation of prior distribution but also on those of the object data for the measurement of information content.

1. ま え が き

水工計画の重要な指標である確率水文量あるいはリターンピリオドの精度は、(1)分布形、(2)パラメタ推定法、(3)データの特性、に依存して決まると筆者らは考えている。

(1)については、1変数の場合多くの分布形が存在する¹⁾が、確率水文量の安定性の立場から見ると、まだ不十分な点があるとともに、2変数以上では分布形そのものの存在も少なくなり、未解決な問題が多い。(2)には、従来主として、モーメント法と最尤法が用いられてきた。一方、1979年、Greenwood ら²⁾は Probability Weighted Moments 法

* 土木工学科 助手

** 土木工学科 教授

*** 矢作建設工業(株)(研究当時、大学院工学研究科土木工学専攻学生)

**** 大学院工学研究科土木工学専攻学生

(PWM法)を提案し、現在竹内ら³⁾は、その有効性を検討しつつある。しかし、この推定法には、まだ今後の研究に委ねなければならない課題が多く残されているようである。(3)は、実験データが得られない水文統計の分野では、古くから大きな関心事であったが、理論的な取り扱いがなされないままになっていた。最近、筆者ら^{4),5),6),7),8),9)}は、これを、時系列的に得られた一連の水文データが、確率分布のパラメータ推定精度をどの程度高めるかという立場から議論した。

ところが、水文データの中には、かなり観測精度の劣るデータが多数含まれている。また、未観測地点のデータを何らかの方法により推定して用いる場合もある。これは、広い意味で水越¹⁰⁾の気候復元も含んでいる。このような時、対象としたデータの採否を決定するには、設定した事前状態に対する個々のデータのもつ情報量と精度について、十分に把握しておかなければならない。

本研究は、上述の目的のために、1986年の筆者ら⁴⁾の研究を理論的背景として、基礎確率変数が正規分布に従う場合に、水文データのもつ情報量の評価法を提案したものであり、その概略は次のとおりである。

第2章では、水文データのもつ情報量を定義し、“平均未知・分散既知”、および“両方とも未知”の2つの場合に分け、さらに、各々の場合について、2つの事前状態に対する情報量算定の理論式を導出する。第3章では、 $N(0, 1^2)$ からシミュレートしたデータと長野県内19ヶ所で観測された年降水データを用いて、第2章の理論式から計算した情報量を示し、それについて考察する。最後に、第4章では、得られた結果を要約して提示するとともに、将来の発展の方向について触れておく。

今後、水文統計は、観測データが十分に、あるいは、正確に得られていない時代に目を向ける必要がある。このとき、このような水文データの情報量を求めることは、水工計画の信頼性を向上させる上で研究しておかなければならない重要な分野であり、本稿は、それに対する基礎的研究に位置付けられる。

2. 情報量の定式化

2.1 水文学量のもつ情報量

一般に、情報とは、“事前の不確定な状態を確定な状態に近づけてくれるもの”、換言すれば、“事前の不確定さを減少させてくれるもの”である。数式的に書けば、

$$[\text{情報量}] = [\text{事前の不確定さ}] - [\text{事後の不確定さ}] \quad (1)$$

となる。

さて、水文統計を対象としたとき、水文学量が従う母集団分布と用いたモデルの確率分布が一致しているなら、推定された確率水文学量の不確定さは、推定母数の不確定さに支配される。従って、このとき、水文データのもつ情報量とは、“事前の推定母数の不確定さの解消に寄与した量”と考えられる。すなわち、(1)式の“不確定さ”を“推定母数の不確定さ”と読み換えればよい。

次に、(1)式を数量化するために、不確定さの尺度である Shannon のエントロピー¹¹⁾の導入を考える。いま、確率変数 x_1, x_2, \dots, x_n が m 個の連続変数で表わされた母数 $\theta_1, \theta_2, \dots, \theta_m$ を持つ同じ確率分布に従うものとする。これらの同時確率密度関数を $\xi(\theta_1,$

$\theta_2, \dots, \theta_m$) とすると, この分布のエントロピーは, 1変数の場合を拡張して, 次のようになる¹²⁾.

$$H(\theta) = - \int \dots \int \xi(\theta_1, \theta_2, \dots, \theta_m) \ln \xi(\theta_1, \theta_2, \dots, \theta_m) d\theta_1 d\theta_2 \dots d\theta_m. \quad (2)$$

上式は, $\xi(\cdot)$ が一様に近い分布からシャープな分布に移行するに従って, 小さな値をとることが知られており, 事前状態の m 個の母数 $\theta_1, \theta_2, \dots, \theta_m$ の不確定さを表現している.

事後状態の不確定さは, 先の事前状態において, $\theta_1, \theta_2, \dots, \theta_m$ の母数をもつ母集団から, x_1, x_2, \dots, x_n の実現値 x'_1, x'_2, \dots, x'_n という水文データを観測したときの事後確率密度関数を $\xi(\theta_1, \theta_2, \dots, \theta_m | x'_1, x'_2, \dots, x'_n)$ とすると, (2)式と同様にして

$$H(\theta | x'_1, x'_2, \dots, x'_n) = - \int \dots \int \xi(\theta_1, \theta_2, \dots, \theta_m | x'_1, x'_2, \dots, x'_n) \cdot \ln \xi(\theta_1, \theta_2, \dots, \theta_m | x'_1, x'_2, \dots, x'_n) d\theta_1 d\theta_2 \dots d\theta_m \quad (3)$$

で表わされる. 従って, (1)式より, 水文データ x'_1, x'_2, \dots, x'_n のもつ情報量 $I(\theta | x'_1, x'_2, \dots, x'_n)$ は, (2)式から(3)式を差し引いて

$$I(\theta | x'_1, x'_2, \dots, x'_n) = H(\theta) - H(\theta | x'_1, x'_2, \dots, x'_n) \quad (4)$$

と記述される.

次に, x_i はデータ発生源の確率変数であるから, その確率密度関数を $p(x_i)$ とするならば, データを得た後の事後状態の平均の不確定さは

$$H(\theta | X) = \int p(x_i) H(\theta | x_i) dx_i \quad (5)$$

となる. 従って, x_i の発生源の情報量は

$$I(\theta | X) = H(\theta) - H(\theta | X) \quad (6)$$

から計算される. ただし, ここで得られた情報量は, 対象とした母数の事前状態に依存するものであるということを十分に認識しておかなければならない.

2.2 事前分布で平均未知・分散既知の場合

前節では, 確率変数 x_i が任意個数の母数 $\theta_1, \theta_2, \dots, \theta_m$ をもつ確率分布 $p(x_i)$ に従う場合について記述したが, 本節と次節では, $p(x_i)$ を確率分布の中で最も基本的な正規分布に限る. このとき, 母数は平均と分散の2個であり, それらをそれぞれ, μ, v と書くことにする. なお, 本節と次節の情報量算定式は, 時系列的な一連の水文データのもつエントロピーを測定した筆者ら⁴⁾の研究を理論的背景として, そこで取り扱われた不確定さの算定式を, 設定した事前状態に対する対象とした個々のデータのもつ情報量を評価する形に変換し, 発展させたものであることを付記しておく.

(1) 事前状態が事前分布そのものの場合

ここでは, 仮定した事前分布から求めた事前状態に, 直接データが入力された場合の情報量を測定する. 水文統計的には, 観測データの全く存在しない地点の統計解析を行いたいとき, 推定データの採用の是非を問う場合を想定している.

いま平均 μ とは無関係に、一定の既知の分散 v をもって、確率変数 x_i が互いに独立に同じ正規分布に従うものとする。次に、 μ を確率変数とみなし、その事前確率分布の密度関数 $\xi(\mu)$ が、平均 m_μ 、分散 v_μ をパラメタとする正規分布で与えられるとする。すなわち

$$\xi(\mu) = f_N(\mu | m_\mu, v_\mu) \quad (7)$$

である。ただし、 $f_N(\cdot)$ は正規分布の密度関数を表わしている。

このとき、 x_1, x_2, \dots, x_n の実現値である n 個のデータ x'_1, x'_2, \dots, x'_n が与えられた後の μ の事後確率分布の密度関数 $\xi(\mu | x'_1, x'_2, \dots, x'_n)$ も、次の正規分布で与えられる。

$$\xi(\mu | x'_1, x'_2, \dots, x'_n) = f_N(\mu | m_{\mu | x'_1, x'_2, \dots, x'_n}, v_{\mu | x'_1, x'_2, \dots, x'_n}) \quad (8)$$

ただし

$$m_{\mu | x'_1, x'_2, \dots, x'_n} = \frac{m_\mu v + n \bar{x} v_\mu}{v + n v_\mu}, \quad (9)$$

$$v_{\mu | x'_1, x'_2, \dots, x'_n} = \frac{v v_\mu}{v + n v_\mu} \quad (10)$$

であり

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x'_i \quad (11)$$

を表わしている。従って、 μ の事前確率分布のエントロピーは、(3)式に(7)式を代入して

$$\begin{aligned} H(M_0) &= - \int_{-\infty}^{\infty} \xi(\mu) \ln \xi(\mu) d\mu \\ &= - \int_{-\infty}^{\infty} \frac{1}{(2\pi v_\mu)^{1/2}} \exp \left\{ -\frac{(\mu - m_\mu)^2}{2v_\mu} \right\} \\ &\quad \cdot \ln \left[\frac{1}{(2\pi v_\mu)^{1/2}} \exp \left\{ -\frac{(\mu - m_\mu)^2}{2v_\mu} \right\} \right] d\mu \\ &= \ln(2\pi e v_\mu)^{1/2} \end{aligned} \quad (12)$$

となり、上式は m_μ を含まないことがわかる。一方、 v_μ は μ の事前分布の分散であるため何らかの方法で仮定しなければならない。ここでは、標本平均に関する中心極限定理の性質を利用して、次のように設定する。

$$v_\mu = \frac{v'}{n_{\max}/2} \quad (13)$$

ただし、 v' は全データ数 n_{\max} 個のデータから算定された分散である。

結局、事前確率分布のエントロピーは(12)式より

$$H(M_0) = \ln\left(2\pi e \frac{v'}{n_{\max}/2}\right)^{1/2} \quad (14)$$

で与えられることになる。

同様に、事後確率分布のエントロピーは、(3)式に(8)式を代入して

$$\begin{aligned} H(M_0|x'_1, x'_2, \dots, x'_n) &= \ln(2\pi e v_{\mu|x'_1, x'_2, \dots, x'_n}) \\ &= \ln\left(2\pi e \frac{v v_{\mu}}{v + n v_{\mu}}\right)^{1/2} \end{aligned} \quad (15)$$

となる。ここで、 v_{μ} は (13) 式で、 v は v' で与えられるため、上式は

$$H(M_0|x'_1, x'_2, \dots, x'_n) = \ln\left(2\pi e \frac{v'}{n_{\max}/2 + n}\right)^{1/2} \quad (16)$$

と書き換えられる。従って、1つのデータ x'_i を得た後の不確定さ、すなわちエントロピーは、上式で $n = 1$ とおいて次式のように得られる。

$$H(M_0|x'_i) = \ln\left(2\pi e \frac{v'}{n_{\max}/2 + 1}\right)^{1/2} \quad (17)$$

以上より、求める情報量は、事前分布のエントロピーとデータを1つ得た後の事後分布のエントロピーとの差であるため、(14)式から(17)式を差し引いて

$$I(M_0|x'_i) = \ln\left(\frac{n_{\max}/2 + 1}{n_{\max}/2}\right)^{1/2} \quad (18)$$

となり、上式は n_{\max} のみに依存していることがわかる。

(2) 事前状態にすべてのデータを考慮した場合

ここで設定した事前状態は、現存する利用可能なデータをすべて考慮しているため、現在の状態にデータが入力された場合の情報量を測定していることになる。すなわち、この状態は現存する利用可能なデータに、精度の劣るデータや推定データを追加することの是非を議論する場合を想定している。このとき、事前状態のエントロピーは、(16)式の n に n_{\max} を代入して

$$H(M_{\max}) = \ln\left(2\pi e \frac{v'}{3n_{\max}/2}\right)^{1/2} \quad (19)$$

となる。

同様に、データ x'_i を1つ得た後の事後状態のエントロピーは、(16)式の n に " $n_{\max} + 1$ " を代入して

$$H(M_{\max}|x'_i) = \ln\left(2\pi e \frac{\nu'}{3n_{\max}/2+1}\right)^{1/2} \quad (20)$$

と書かれる。

よって、対象としたデータ x'_i がもつ情報量は、(19)式から(20)式を差し引いた

$$I(M_{\max}|x'_i) = \ln\left(\frac{3n_{\max}/2+1}{3n_{\max}/2}\right)^{1/2} \quad (21)$$

から算定される。上式も(18)式と同様 n_{\max} のみに依存していることがわかる。

2.3 事前分布で平均・分散共に未知の場合

(1) 事前状態が事前分布そのものの場合

ここでは、2.2(1)と同様にして、仮定した事前分布から求めた事前状態にデータが入力された場合の情報量を測定する。

いま、 x_i は互いに独立に平均 μ 、精度 $h \equiv 1/\nu$ をもって正規分布に従うものとする。次に、 μ と h の両者を確率変数とみなし、 (μ, h) の同時事前確率分布の密度関数 $\xi(\theta, h)$ が、 $\alpha', \beta', n', \nu'$ をパラメタとする次の正規-ガンマ分布で与えられるとする。

$$\begin{aligned} \xi(\mu, h) &= \frac{n'^{1/2}}{(2\pi)^{1/2}} \frac{1}{\Gamma(\nu'/2)} \left(\frac{1}{2} \beta' \nu'\right)^{\nu'/2} \exp\left\{-\frac{1}{2} n' h (\theta - \alpha')^2\right\} \\ &\quad \cdot \exp\left(-\frac{1}{2} \beta' \nu' h\right) h^{(\nu'-1)/2} \end{aligned} \quad (22)$$

ここに $\Gamma(\cdot)$ はガンマ関数である。

このとき

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x'_i, \quad S = \frac{1}{n-1} \sum_{i=1}^n (x'_i - \bar{x})^2 \quad (23)$$

で算定される (μ, h) に関する十分統計量が得られた後の (μ, h) の同時事後分布の密度関数 $\xi(\mu, h|x'_1, x'_2, \dots, x'_n)$ もやはり次の正規-ガンマ分布で与えられる。

$$\begin{aligned} \xi(\mu, h|x'_1, x'_2, \dots, x'_n) &= \frac{n''^{1/2}}{(2\pi)^{1/2}} \frac{1}{\Gamma(\nu''/2)} \left(\frac{1}{2} \beta'' \nu''\right)^{\nu''/2} \exp\left\{-\frac{1}{2} n'' h (\mu - \alpha'')^2\right\} \\ &\quad \cdot \exp\left(-\frac{1}{2} \beta'' \nu'' h\right) h^{(\nu''-1)/2} \end{aligned} \quad (24)$$

ただし

$$n'' = n' + n, \quad (25)$$

$$\alpha'' = \frac{n' \alpha' + n \bar{x}}{n' + n}, \quad (26)$$

$$\beta'' = \frac{(\nu' \beta' + n' \alpha'^2) + (\nu S + n \bar{x}^2) - n'' \alpha''^2}{\nu' + n}, \quad (27)$$

$$\nu'' = \nu' + n \quad (28)$$

であり、ここで

$$\nu = n - 1 \quad (29)$$

と置いている。

従って、 (μ, h) の事前分布のエントロピーは、(3)式に(22)式を代入して次式から計算される。

$$\begin{aligned} H(M_0) &= - \int_0^\infty \int_{-\infty}^\infty \xi(\mu, h) \ln \xi(\mu, h) d\mu dh \\ &= - \int_0^\infty \int_{-\infty}^\infty \frac{n'^{1/2}}{(2\pi)^{1/2}} \frac{1}{\Gamma(\nu'/2)} \left(\frac{1}{2} \beta' \nu'\right)^{\nu'/2} \cdot \exp\left\{-\frac{1}{2} n' h (\mu - \alpha')^2\right\} \\ &\quad \cdot \exp\left(-\frac{1}{2} \beta' \nu' h\right) \cdot h^{(\nu'-1)/2} \ln \left[\frac{n'^{1/2}}{(2\pi)^{\nu'/2}} \cdot \frac{1}{\Gamma(\nu'/2)} \cdot \left(\frac{1}{2} \beta' \nu'\right)^{\nu'/2} \right. \\ &\quad \left. \cdot \exp\left\{-\frac{1}{2} n' h (\mu - \alpha')^2\right\} \cdot \exp\left(-\frac{1}{2} \beta' \nu' h\right) \cdot h^{(\nu'-1)/2} \right] d\mu dh \\ &= - \ln A_1 + B_1 \left\{ 1 - \Psi\left(B_1 + \frac{1}{2}\right) + \ln C_1 \right\} + 1 \end{aligned} \quad (30)$$

ただし

$$A_1 = \frac{n'^{1/2}}{(2\pi)^{1/2}} \frac{1}{\Gamma(\nu'/2)} \cdot \left(\frac{1}{2} \beta' \nu'\right)^{\nu'/2}, \quad (31)$$

$$B_1 = \frac{\nu' - 1}{2}, \quad (32)$$

$$C_1 = \frac{\beta' \nu'}{2} \quad (33)$$

であり、 $\Psi(\cdot)$ はディ・ガンマ関数を表現している。

同様に、事後確率分布のエントロピーは、(3)式に(24)式を代入して

$$H(M_0 | x'_1, x'_2, \dots, x'_n) = - \ln A_2 + B_2 \left\{ 1 - \Psi\left(B_2 + \frac{1}{2}\right) + \ln C_2 \right\} + 1 \quad (34)$$

となる。ただし

$$A_2 = \frac{n''^{1/2}}{(2\pi)^{1/2}} \frac{1}{\Gamma(\nu''/2)} \left(\frac{1}{2} \beta'' \nu''\right)^{\nu''/2}, \quad (35)$$

$$B_2 = \frac{\nu'' - 1}{2}, \quad (36)$$

$$C_2 = \frac{\beta'' \nu''}{2} \quad (37)$$

である。

以上より、求める情報量は、(30)式から(34)式を差し引いて次式から算定される。

$$\begin{aligned} I(M_0|x'_1, x'_2, \dots, x'_n) &= \ln \frac{A_2}{A_1} + B_1 \left\{ 1 - \Psi \left(B_1 + \frac{1}{2} \right) + \ln C_1 \right\} \\ &\quad - B_2 \left\{ 1 - \Psi \left(B_2 + \frac{1}{2} \right) + \ln C_2 \right\} \end{aligned} \quad (38)$$

ここで、事前分布のパラメタは、それらがもつ統計的性質から次のように与えることにする。

$$n' = n_{\max}, \quad (39)$$

$$\alpha' = \frac{1}{n'} \sum_{i=1}^{n'} x'_i, \quad (40)$$

$$\beta' = \frac{1}{n' - 1} \sum_{i=1}^{n'} (x'_i - \alpha')^2, \quad (41)$$

$$\nu' = n' - 1 \quad (42)$$

また、事後分布のパラメタ n'' , α'' , ν'' は、1個のデータがもつ情報量を測定するとき、 n は1となり、(23)式の \bar{x} は x'_i に、 S は零になるため、(25)~(28)式は次のように書き換えて、求められる。

$$n'' = n' + 1, \quad (43)$$

$$\alpha'' = \frac{n' \alpha' + x'_i}{n' + 1}, \quad (44)$$

$$\beta'' = \frac{(\nu' \beta' + n' \alpha'^2) + x'_i{}^2 - n' \alpha'^2}{\nu' + 1}, \quad (45)$$

$$\nu'' = \nu' + 1. \quad (46)$$

(2) 事前状態にすべてのデータを考慮した場合

この場合の1つのデータ x'_i の持つ情報量は、(1)と同様にして、次のように書かれる。

$$\begin{aligned} I(M_{\max}|x'_i) &= \ln \frac{A_2}{A_1} + B_1 \left\{ 1 - \Psi \left(B_1 + \frac{1}{2} \right) + \ln C_1 \right\} \\ &\quad - B_2 \left\{ 1 - \Psi \left(B_2 + \frac{1}{2} \right) + \ln C_2 \right\} \end{aligned} \quad (47)$$

ただし

$$A_1 = \frac{n''_1}{(2\pi)^{1/2}} \cdot \frac{1}{\Gamma(\nu''_1/2)} \cdot \left(\frac{1}{2}\beta''_1 \nu''_1\right)^{\nu''_1/2}, \quad (48)$$

$$A_2 = \frac{n''_2}{(2\pi)^{1/2}} \cdot \frac{1}{\Gamma(\nu''_2/2)} \cdot \left(\frac{1}{2}\beta''_2 \nu''_2\right)^{\nu''_2/2}, \quad (49)$$

$$B_1 = \frac{1}{2}(\nu''_1 - 1), \quad (50)$$

$$B_2 = \frac{1}{2}(\nu''_2 - 1), \quad (51)$$

$$C_1 = \frac{1}{2}\beta''_1 \nu''_1, \quad (52)$$

$$C_2 = \frac{1}{2}\beta''_2 \nu''_2 \quad (53)$$

であり、また

$$n''_1 = 2n', \quad (54)$$

$$n''_2 = 2n' + 1, \quad (55)$$

$$\alpha''_1 = \frac{n'\alpha' + n'\bar{x}_1}{2n'}, \quad (56)$$

$$\alpha''_2 = \frac{n'\alpha' + (n' + 1)\bar{x}_2}{2n' + 1}, \quad (57)$$

$$\beta''_1 = \frac{(\nu'\beta' + n'\alpha'^2) + (\nu'S_1 + n'\bar{x}_1) - n''_1\alpha''_1{}^2}{\nu' + n'}, \quad (58)$$

$$\beta''_2 = \frac{(\nu'\beta' + n'\alpha'^2) + \{(\nu' + 1)S_2 + (n' + 1)\bar{x}_2\} - n''_2\alpha''_2{}^2}{\nu' + n' + 1}, \quad (59)$$

$$\nu''_1 = \nu' + n', \quad (60)$$

$$\nu''_2 = \nu' + n' + 1 \quad (61)$$

である。ここで、下付き“1”は全データ数から設定した事前状態を、“2”はその事前状態に情報量を測定するデータ x'_i を追加した事後状態を表わしている。従って、 \bar{x}_1 、 S_1 は n_{\max} 個のデータを用いて、 \bar{x}_2 、 S_2 は $(n_{\max} + 1)$ 個のデータを用いて(23)式から計算された値である。また、下付きのない文字は、(1)で計算された値を用いることを意味している。

3. データへの適用と考察

3.1 シミュレーションデータの場合 (平均・分散共に未知の場合)

前章で提示された理論は、 $p(x_i)$ が正規分布に従う場合、すなわち正規情報が得られる場合であった。従って、シミュレーションデータとして $N(0, 1^2)$ から発生された乱数を用い、実データへの適用に先がけて規準化された正規乱数をもつ情報量を測定することにより、(38)式および(47)式の基礎特性を把握しておく。なお、平均だけが未知の場合は、(18)式および(21)式で示したように、情報量はデータ数 n_{\max} だけに依存して決まるため、ここでは省略した。

図一1は、事前状態が仮定した事前分布そのもの場合、すなわち、(38)式から計算された100個の正規乱数をもつ情報量である。この図より、平均付近のデータのもつ情報量は小さく、平均より離れる程2次関数的に大きくなること、および、得られた情報量を結ぶ曲線はデータの平均値を通る縦軸に対して左右対称形になっていることがわかる。

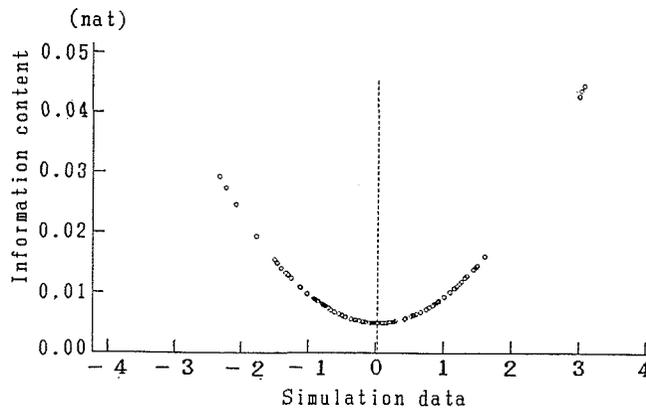


Fig. 1 Information content given by simulation data (mean: variate, variance: variate, $(I(M_0|x_i))$).

図一2は、図一1と同じ乱数を用いて、事前状態にすべてのデータを考慮した場合、すなわち、(47)式から計算した情報量を示したものである。この場合、個々のデータのもつ情報量の特徴は図一1と同様であるが、同一のデータについて、その値を比較すると、どのデータについても、図一2の方が小さくなっている。このことは、事前状態ですべてのデータを考慮することにより、事前の母数推定精度が高い状態に改善されているため、同じデータでも事前状態に対する影響力が小さくなったために生じたものと推察され、情報量を(1)式で定義したことからして当然の帰結と言えよう。

3.2 実データの場合

水流量は、一般に対象とする時間単位に依存して従う分布形が変化し、時間単位が大きい程正規分布に近づくことが知られている¹³⁾。従って、ここでは水文データが正規分布に

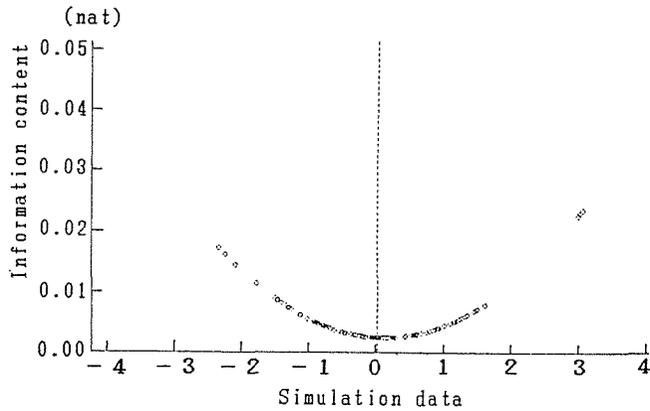


Fig. 2 Information content given by simulation data (mean: variate, variance: variate, $(I(M_{\max}|x_i))$).

従う典型的な場合として、長野県内19ヶ所で観測された年降水量を用いることにする¹⁴⁾。それらの観測所名とデータ数は表-1のとおりであり、データ数の最少は宗賀の41個、最多は長野の94個である。

Table 1 Rain gauge stations and their numbers of data.

No.	Rain gauge stations	Number of Data
1	Azumi	5 6
2	Bessho-Onsen	5 1
3	Iiyama	6 8
4	Iwamura	6 1
5	Kinasa	6 0
6	Kitamaki	4 7
7	Matsumoto	8 5
8	Soga	4 1
9	Nagano	9 4
10	Nakajo	5 5
11	Nakano	7 6
12	Nozawa-Onsen	6 5
13	Oomati	8 0
14	Shinonoi	6 7
15	Suzaka	7 2
16	Toyoshina	5 5
17	Ueda	8 8
18	Wada	5 1
19	Yashiro	6 4

(1) 事前状態で平均未知・分散既知の場合

平均だけが未知の場合は、3.1でも触れたように、データのもつ情報量はその観測所のデータ数だけに依存して決まり、データの値とは無関係に一定である。すなわち、正規分布の平均推定という立場からみると、どのデータも同量の情報を有していることになる。図-3は、事前状態が仮定した事前分布そのものの場合、すなわち、(18)式の計算結果を観測所ごとに示したものである。事前分布のパラメタの与え方、すなわち、(13)式の値の

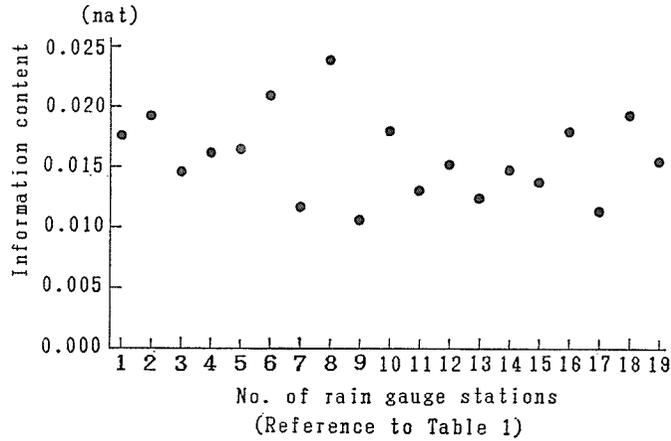


Fig. 3 Information content given by annual rainfall (mean: variate, variance:known, $(I(M_o|x_i))$).

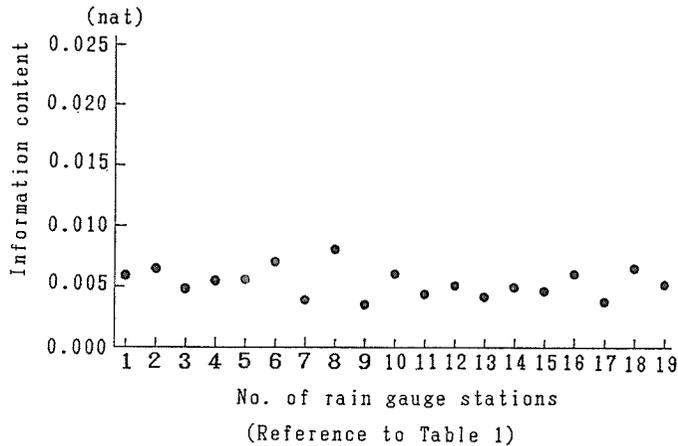


Fig. 4 Information content given by annual rainfall (mean: variate, variance:known, $(I(M_{\max}|x_i))$).

相違により観測所間の情報量の大きさに差異が存在するが、当然1つの観測所内ではどのデータも同じ情報量をもっていることがわかる。

図-4は、事前状態にすべてのデータを考慮した場合、すなわち(21)式の計算結果を示したものである。図-3と比較して、観測所ごとの情報量の大きさの順位には変動はないが、値自身はかなり小さくなっている。これは、前節と同様に、事前の母数推定精度が高い状態に改善されているためと推察される。また、観測所ごとの情報量の相違は、(13)式の $v\mu$ の値にも依存するが、事前状態を形成するときに考慮したデータ数に支配される。すなわち、データ数が多い観測所程、換言すれば、事前の情報が多い観測所程、対象とし

たデータのもつ情報量は小さくなり、長野は19の観測所の中で最小の情報量を示していることがわかる。

(2) 事前分布で平均・分散共に未知の場合

ここでは、シミュレーションデータに対する前節の計算を、(1)で用いた年降水量に対して行う。

図一5は図一1に対応した、事前状態が仮定した事前分布そのもので与えられる場合であり、(38)式から計算された情報量の1例として、長野、松本、上田について示したものである。他の16地点についても、データの平均値および情報量を連ねた曲線の曲率には差異はあるものの、同様の挙動を示している。観測所ごとの相違は、(39)~(42)式の値を用い

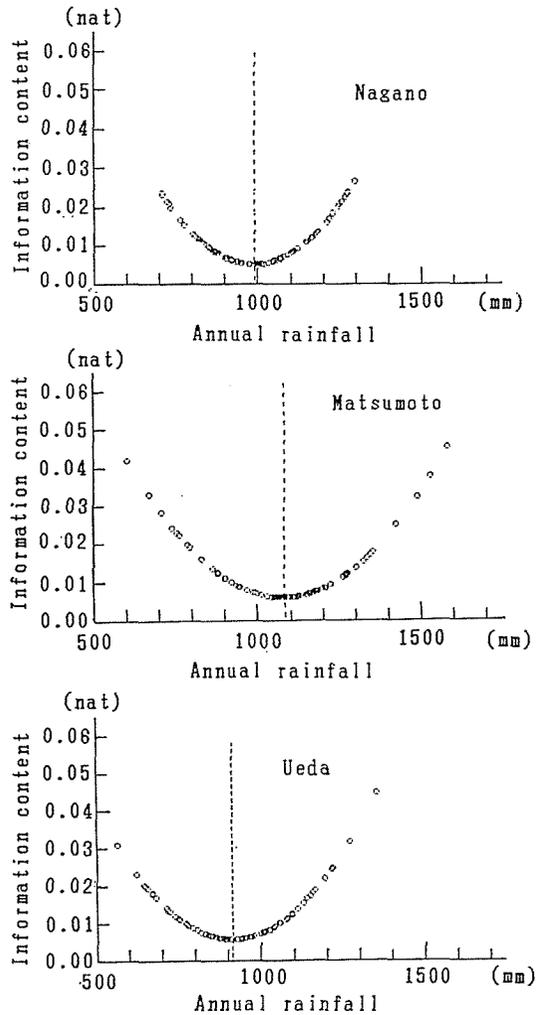


Fig. 5 Information content given by annual rainfall (mean: variate, variance: variate, $(I(M_0|x_i))$).

て計算された事前状態と対象としたデータ自身の値に依存して生じており、データ数だけで論じることはできない。しかしながら、どの観測所についても、図一1で考察したのと同じように、平均付近の水文データは、たびたび観測されるため情報量が小さいのに対し、平均から大きく離れて、めったに観測されない水文データ程、情報量の値は2次関数的に大きくなり、母数の推定精度を高めることになっている。また、得られた情報量を結ぶ曲線は、データの平均値を通る縦軸に対して左右対称形になっている。このことは、水文データがもつ情報という概念に関する我々の直感に一致している。

図一6は図一2に対応した、事前状態にすべてのデータを考慮した場合であり、(47)式から計算された情報量の1例として、長野、松本、上田について示したものである。他の

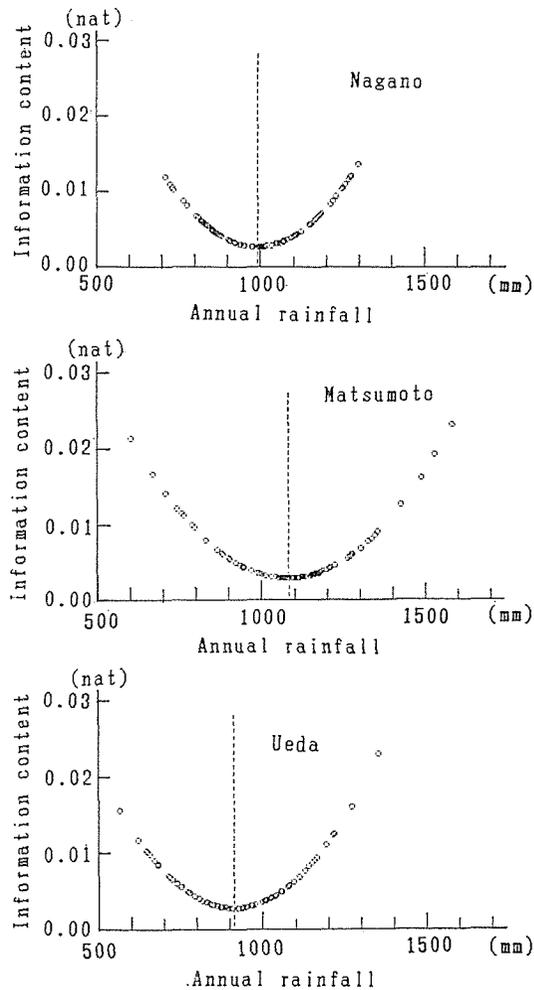


Fig. 6 Information content given by annual rainfall (mean: variate, variance: variate, $(I (M_{\max} | x_i))$).

16地点も同様の挙動を示している。また、個々のデータについて図-5と図-6を比較すると、図-6の方がかなり小さな情報量となっている。これは、3.1で述べた理由と同様に、図-6の場合の方が事前状態が大きく改善されているためと推察される。

4. 結 論

本論文では、水文データ数の増加をはかるために、従来ほとんど定量的に取り扱われていなかった水文データのもつ情報量の測定法を提案したものである。

基礎確率変数が正規分布に従う場合に、 $N(0, 1^2)$ からのシミュレーションデータと年降水量を用いた計算から得られた結果を要約して提示すると、次のようになる。

1) 事前分布で平均未知・分散既知、すなわち平均のみを推定するという立場では、データのもつ情報量はデータの値に依存せず一定であり、観測所のデータ数のみの関数として表現される。

2) 事前分布で平均・分散共に未知の場合、すなわち両母数を同時に推定するという立場でのデータのもつ情報量は、平均値付近では小さく、平均値から離れる程2次関数的に大きくなり、また平均値を通る縦軸を中心に左右対称となる。これは、水文データがもつ情報という概念に関する我々の直感に一致する。

3) 上2者のどちらの場合も、対象とするデータを入力する前の状態、すなわち事前状態の設定の仕方により、同じデータでも情報量の値は異なり、すべてのデータを考慮した事前状態の方が仮定した事前分布そのもので与えた事前状態より小さな情報量をもつ。これは、すべてのデータを組み入れることにより、パラメタ推定精度の高い状況を事前状態として設定したためと推察され、本稿では不確定さの減少度として情報量を定義していることからみて、当然の帰結と言えよう。

4) 観測所ごとの情報量の相違は、1)の場合は、データを考慮しない事前状態でのエントロピー $H(M_0)$ を計算するときの事前分布のパラメタ設定法、事前状態に考慮したデータ数、2)の場合は、上の2つの場合に加えて、事前状態に考慮したデータの特性と情報量を測定するデータの特性によって生じ、同じ大きさのデータでも観測所ごとにそのデータのもつ情報量の値が異なることになる。

本稿は、水文データのもつ情報量測定の初めての試みであったため、基礎確率変数が正規分布に従う場合、すなわち年降水量、年流量等の時間単位の長い水文量に対する理論式を提示して議論したが、将来、基礎確率変動の従う分布形を、指数分布、対数正規分布、ガンマ分布、ガンベル分布、最大エントロピー分布等に拡張していくことにより、種々の時間単位の水文データの情報量を測定していきたいと考えている。さらにいま、水文データの精度の定量化をはかることにより、上述の情報量との関連で対象としたデータ採否のボーダーラインを求める研究へと本稿を進展させることを検討している。

参 考 文 献

- 1) 例えば、神田 徹，藤田睦博：水文学——確率論的手法とその応用——；pp.31-47，技報堂出版，(1982年)。

- 2) Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. : Probabiliy Weighted Moments : Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form, Water Resour. Res., Vol. 15, (No.5), pp.1049-1054 (1979).
- 3) 竹内邦良, 土屋一仁 : 正規分布のパラメータの PWM 解, 水理講演会論文集, 第31回, pp.191-196 (1987年).
- 4) 寒川典昭, 荒木正夫, 渡辺輝彦 : 確率分布の推定母数の不確定性評価法, 土木学会論文集, II-6, pp.133-141 (1986年).
- 5) 荒木正夫, 寒川典昭, 上原 剛 : 指数分布・対数正規分布の母数推定の信頼性, 土木学会中部支部研究発表会講演概要集, II-3, pp.76-77 (1986年).
- 6) 上原 剛, 荒木正夫, 寒川典昭 : 非正規母集団の推定母数の信頼性 (その2), 土木学会年次学術講演会講演概要集, II-47, pp.93-94 (1986年).
- 7) 荒木正夫, 寒川典昭, 上原 剛, 草刈智一 : 確率水文学の信頼性について, 自然災害科学中部地区シンポジウム講演要旨集, pp.6-7 (1986年).
- 8) 荒木正夫, 寒川典昭, 上原 剛, 草刈智一 : 非正規母集団の推定母数の信頼性 (その3), 土木学会中部支部研究発表会講演概要集, II-2, pp.104-105 (1987年).
- 9) 草刈智一, 寒川典昭, 荒木正夫, 上原 剛 : ガンベル分布の推定母数の信頼性, 土木学会年次学術講演会講演概要集, II-1, pp.32-33 (1987年).
- 10) 水越允治 : 近畿・東海地方における近世の気候復元——とくに乾湿条件について——, 京都大学防災研究所年報, 第28号, B-2, pp.121-132 (1985年).
- 11) 例え, 笠原芳郎 : 情報理論と通信方式, p.150, 共立出版, 1965年.
- 12) 11) に同じ, p.161.
- 13) 角屋 睦 : 水文統計論, 水工学シリーズ, 64-02, p.2 (1964年).
- 14) 建設省北陸地方建設局千曲川工事事務所編集 : 千曲川・犀川河川要覧 (雨量資料編), 1976年.