

2母数対数正規分布の推定母数の信頼性評価

Reliability Evaluation on Estimated Parameters of 2-Parameter Lognormal Distribution

信州大学工学部

寒川 典昭 Noriaki SOGAWA

信州大学工学部

荒木 正夫 Masao ARAKI

長野県庁

上原 剛 Tsyoshi UEHARA

中央復建コンサルタンツ(株)

草刈 智一 Tomokazu KUSAKARI

The reliability of the estimated parameters is evaluated by using the entropy to posterior distribution in the 2-parameter lognormal distribution. When λ and ζ are parameters representing the expectation and the variance of the hydrological value transformed into logarithm respectively, the following cases are discussed. ① λ :unknown, ζ :known, ② λ :known, ζ :unknown, ③ λ, ζ :unknown. In each case, the posterior distribution of the unknown parameter(s) is derived from Bayes' Theorem, and its entropy is developed. This model is applied to monthly rainfall, and the increase of the reliability, accompanied by the increase of the number of data, is investigated.

Keywords : lognormal distribution, parameters, reliability, entropy, monthly rainfall

1. まえがき

水文データは一般に小標本であるため、特にリターンピリオドの大きい確率水量は不確定であることが知られている^{1), 2), 3)}。著者らは、水量に母集団分布が存在し、それに定常性が仮定できる場合、この不確定さを支配する要因の主なもの、確率分布の選択の問題、母数推定法の選択の問題、及びデータの不十分さの問題にあると考えている³⁾。前2者は、従来から確率分布を用いた水文統計の中心課題であり、多くの研究成果が発表されているとともに、その適用比較も行なわれてきている^{4), 5), 6), 7)}。一方、3番目の問題、具体的にはデータ数と、確率水量あるいはリターンピリオドの信頼性との関係を明らかにしようとする研究は、長い間本格的な取り組みがなされないままになっていた。その原因の一つはこの問題の定式化の困難さにあった。

このような背景を踏まえて、寒川らは3番目の問題へのアプローチを試みた。寒川らの研究は、年降水量と年最大日降水量のデータを用いて求めた確率分布と確率水量の変動をデータとの関連で明らかにした上で³⁾、正規分布³⁾とガンベル分布⁸⁾の母数推定の精度がデータ数の増加とともにどのように向上するかを、母数の事後分布のエントロピーから評価したものである。

ところが、水量が上述の確率分布に従わないことが多々ある。例えば、渇水生起の確率評価等を議論するとき、旬から月程度の時間単位の降水・流量データを必要とするが、この場合これらの水量は対数正規分布を示すことが知られている。このように、対数正規分布は水文統計上重要な役割を演じる分布の1つである。従って本稿では、上述の研究を2母数対数正規分布に拡張し、データ数の増加と推定母数の信頼性の関係をエントロピー的立場から把握するための理論式を提案し、月降水データへの適用を通じて本法の有効性を検証する。

なお、高棹・池淵⁹⁾は、降水観測がもたらす情報量を確率分布の母数のエントロピーから測定しており、これが本研究の考え方の基礎になっていることを付言しておく。

2. 理論式

ここでは、推定母数の信頼性を評価するために、分布形を支配する母数を確率変数と考え¹⁰⁾、ベイズ論

的に求めた母数の事後確率分布のエントロピーを計算可能な形に変換する。基礎確率変数（以下では母数も確率変数とみなしているため、それと区別する意味で水文量を表現する確率変数を基礎確率変数と呼ぶ。第3章では、これは月降水量を表現することになる。） x が従う分布形として、次の2母数対数正規分布を用いる。

$$f(x) = \frac{1}{\sqrt{2\pi} \xi x} \exp \left\{ -\frac{(\ln x - \lambda)^2}{2\xi} \right\} \quad (1)$$

ここに、 λ 、 ξ は母数である。

2.1 λ を未知、 ξ を既知とした場合

いま、 n 個の基礎確率変数 $x_1, x_2, \dots, x_n$ が、未知母数 λ 、及び λ と無関係な一定の既知母数 ξ をもって、互いに独立に(1)式のガンベル分布に従うものとする。 λ の事前分布に、共役事前確率分布である次の正規分布を仮定する。

$$g_1(\lambda) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(\lambda - \mu)^2}{2\sigma} \right\} \quad (2)$$

ここに、 μ 、 σ は λ に関する平均、分散を表わす母数である。このとき、 $x_1, x_2, \dots, x_n$ の実現値 $x_1', x_2', \dots, x_n'$ が得られた後の λ の事後分布は

$$g_1(\lambda | x_1', x_2', \dots, x_n') = \frac{1}{\sqrt{2\pi}\sigma'} \exp \left\{ -\frac{(\lambda - \mu')^2}{2\sigma'} \right\} \quad (3)$$

なる正規分布となる。ここに、

$$\mu' = \frac{\xi \mu + n \sigma (\overline{\ln x})}{\xi + n \sigma}, \quad (4)$$

$$\sigma' = \frac{\xi \sigma}{\xi + n \sigma} \quad (5)$$

であり、 $\overline{\ln x}$ は $\ln x_i$ ($i=1, 2, \dots, n$)の期待値を意味している。

従って、(3)式から λ の事後分布のエントロピーは次式で表わされる。

$$\begin{aligned} H(\lambda) &= - \int_{-\infty}^{\infty} g_1(\lambda | x_1', x_2', \dots, x_n') \cdot \ln g_1(\lambda | x_1', x_2', \dots, x_n') d\lambda \\ &= \ln(2\pi e \sigma')^{1/2} \end{aligned} \quad (6)$$

2.2 λ を既知、 ξ を未知とした場合

ここでは、2.1の λ と ξ を置き換えればよい。ただし、 ξ の共役事前確率分布が求められていないため、その事前分布に $[\xi_1, \xi_2]$ の区間における次の一様分布を仮定する。

$$g_2(\xi) = \frac{1}{\xi_2 - \xi_1} \quad (7)$$

このとき、 $x_1, x_2, \dots, x_n$ の実現値 $x_1', x_2', \dots, x_n'$ が得られた後の ξ の事後分布は

$$g_2(\xi | x_1', x_2', \dots, x_n') = K_1 \frac{1}{\xi^B} \exp\left(-\frac{A}{\xi}\right) \quad (8)$$

となる。ここに、

$$A = -\left\{ \sum_{i=1}^n (\ln x_i)^2 - 2\lambda \sum_{i=1}^n \ln x_i + n\lambda^2 \right\} / 2, \quad (9)$$

$$B = n/2, \quad (10)$$

$$K_1 = 1 / \left[\frac{e^{At}}{A} \sum_{r=0}^{B-2} (-1)^r \frac{(B-2)! t^{B-r-2}}{(B-r-2)! r!} \right]_{t=1/\xi_1}^{t=1/\xi_2} \quad (11)$$

である。

従って、(8) 式からこの事後分布のエントロピーは次式で表わされる。

$$H(\xi) = - \int_{\xi_1}^{\xi_2} g_2(\xi | x_1', x_2', \dots, x_n') \cdot \ln g_2(\xi | x_1', x_2', \dots, x_n') d\xi$$

$$= -\ln K_1 + K_1 B \int_{\xi_1}^{\xi_2} \frac{1}{\xi^B} \exp\left(\frac{A}{\xi}\right) \ln \xi d\xi + \left[\frac{e^{At}}{A} \sum_{r=0}^{B-1} (-1)^r \frac{(B-1)! t^{B-r-1}}{(B-r-1)! A^r} \right]_{t=1/\xi_2}^{t=1/\xi_1} \quad (12)$$

2.3 λ, ξ 共に未知の場合

ここでは、基礎確率変数 $x_1, x_2, \dots, x_n$ が、2つの無関係な未知母数 λ, ξ をもって、互いに独立に(1) 式の対数正規分布に従うものとする。 $[\lambda_1, \lambda_2], [\xi_1, \xi_2]$ の同時事前分布に

$$g_3(\lambda, \xi) = \frac{1}{(\lambda_2 - \lambda_1)(\xi_2 - \xi_1)} \quad (13)$$

なる一様分布を仮定すると、 λ, ξ の同時事後分布は

$$g_3(\lambda, \xi | x_1', x_2', \dots, x_n') = K_2 \frac{1}{\xi^{n/2}} \exp\left\{-\frac{1}{2\xi}(C-2\lambda D + n\lambda^2)\right\} \quad (14)$$

となる。ここに、

$$C = \sum_{i=1}^n (\ln x_i)^2, \quad (15)$$

$$D = \sum_{i=1}^n \ln x_i, \quad (16)$$

$$K_2 = 1 / \int_{\lambda_1}^{\lambda_2} \int_{\xi_1}^{\xi_2} \frac{1}{\xi^{n/2}} \exp\left\{-\frac{1}{2\xi}(C-2\lambda D + n\lambda^2)\right\} d\xi d\lambda \quad (17)$$

である。

従って、(14)式から λ, ξ の同時事後分布のエントロピーは次式で表わされる。

$$H(\lambda, \xi) = - \int_{\lambda_1}^{\lambda_2} \int_{\xi_1}^{\xi_2} g_3(\lambda, \xi | x_1', x_2', \dots, x_n') \cdot \ln g_3(\lambda, \xi | x_1', x_2', \dots, x_n') d\xi d\lambda$$

$$= -\ln K_2$$

$$- \frac{K_2 n}{2} \int_{\lambda_1}^{\lambda_2} \int_{\xi_1}^{\xi_2} \ln\left(\frac{1}{\xi}\right) S(\lambda, \xi) d\xi d\lambda$$

$$+ \frac{K_2 C}{2} \int_{\lambda_1}^{\lambda_2} \int_{\xi_1}^{\xi_2} \frac{1}{\xi} S(\lambda, \xi) d\xi d\lambda$$

$$- K_2 D \int_{\lambda_1}^{\lambda_2} \int_{\xi_1}^{\xi_2} \frac{\lambda}{\xi} S(\lambda, \xi) d\xi d\lambda$$

$$+ \frac{K_2 n}{2} \int_{\lambda_1}^{\lambda_2} \int_{\xi_1}^{\xi_2} \frac{\lambda^2}{\xi} S(\lambda, \xi) d\xi d\lambda \quad (18)$$

ここに

$$S(\lambda, \xi) = \frac{1}{\xi^{n/2}} \exp\left\{-\frac{1}{2\xi}(C-2\lambda D + n\lambda^2)\right\} d\xi d\lambda \quad (19)$$

である。

3. 月降水量への適用例と考察

ここでは、月降水量データを用いて、前章で計算可能な形に変換した事後分布のエントロピー式を数値計算し、データ数の増加とエントロピーとの関連性を把握する。なお、データとして、千曲川流域内にある長野、松本、上田の6月降水量及び9月降水量を用いた。データ数は6月、9月共に、長野は96個、松本、上

田は84個である。また、データ数は観測した年代の新しいものから順番に1個ずつ増加させた。

3.1 λ を未知、 ξ を既知とした場合

λ の事後分布のエントロピーは(6)式で与えられる。ここで、事前情報である σ' は共役事前確率分布の導出過程を参照して次式のように設定した。

$$\sigma' = \frac{\ln(v/\theta^2 + 1)}{N/2} \quad (20)$$

ここに、 θ 、 v は全データから計算されたデータの平均と分散であり、 N は全データ数である。

図-1の上段はこの場合のデータ数の増加に伴うエントロピーの挙動を示した1例である。この図よりデータ数の増加とともにエントロピーの値は指数関数的に単調減少し、データの値自身には依存していないことがわかる。このことは、(6)式と(20)式からも理解できる。

3.2 λ を既知、 ξ を未知とした場合

ξ の事後分布のエントロピーは(12)式で与えられる。ここで、事前情報である ξ_1 、 ξ_2 、及び既知とした λ は次のように設定した。

1)全データを観測した年代の新しいデータから順番に10個ずつの組に分ける(端数は切り捨てる)。

2)10個ずつに分けた各々の組に対し積率法を用いて ξ の推定値を出し、その中の最小値を ξ_1 、最大値を ξ_2 とする。

3)既知とした λ は、全データに対して積率法を用いて算定する。

表-1に実際に求めた ξ_1 、 ξ_2 と、全データを用いて求めた ξ の推定値を示している。 ξ_2 と ξ_1 の差は、最大で0.3471、最小で0.1327である。

図-1の中段は表-1の ξ_1 、 ξ_2 を用いて計算したデータ数の増加に伴う ξ のエントロピーの挙動を示した1例である。ただし、(11)式において $B=n/2$ 、 $B-2 \geq 0$ であるから、 n を4以上の偶数に限った。計算結果より以下のことが読み取れる。

1)エントロピーの減少の仕方は観測地点及び月によって異なる。

2)エントロピーは増加・減少を繰り返しているが、全体的にみればデータ数の増加とともに減少している。ただし、松本6月の場合は例外的である。

3)長野9月において、データ数が89個から90個になったとき大きくエントロピーが減少しているが、その他の場合データ数が60個程度を越えるとエントロピーの変動は小さくなる。

3.3 λ 、 ξ 共に未知の場合

λ 、 ξ の同時事後分布のエントロピーは(18)式で与えられる。ここで、事前情報である ξ_1 、 ξ_2 は表-1の値を用い、 λ_1 、 λ_2 は3.2で λ と ξ を置き換えて求めた表-2の値を採用した。 λ_2 と λ_1 の差は最大

表-1 事前情報 ξ_1 、 ξ_2 、及び全データを用いて積率法より求めた ξ

Rainfall	ξ_1	ξ_2	ξ
Nagano, Jun.	0.0918	0.4389	0.1929
Nagano, Sept.	0.1685	0.3012	0.2184
Matumoto, Jun.	0.0911	0.3012	0.2263
Matumoto, Sept.	0.1714	0.3364	0.2519
Ueda, Jun.	0.0420	0.2507	0.1868
Ueda, Sept.	0.1656	0.3491	0.2698

表-2 事前情報 λ_1 、 λ_2 、及び全データを用いて積率法より求めた λ

Rainfall	λ_1	λ_2	λ
Nagano, Jun.	4.3284	4.8581	4.6736
Nagano, Sept.	4.6056	5.0051	4.7555
Matumoto, Jun.	4.6203	5.1472	4.8899
Matumoto, Sept.	4.6236	5.1168	4.8898
Ueda, Jun.	4.5461	5.0197	4.7742
Ueda, Sept.	4.4400	5.0646	4.7660

で0.6246, 最小で0.3995である。

図-1の下段は表-1の ξ_1 , ξ_2 , 表-2の λ_1 , λ_2 を用いて計算したデータ数の増加に伴う λ , ξ のエントロピーの挙動を示した1例である。なお, この図には採用した降水量も併記している。当然, 上段, 中段の図にも同じ降水量を用いている。この場合のエントロピーの挙動は3.2の場合と似ており, 長野9月以外のところではデータ数が60個程度を超えるとエントロピーの変動は小さくなり, この数程度以降での推定母数の信頼性には大きな変化がみられないことがわかる。

4. 主要な結論

確率水文学の精度を支配する主要な要因の1つは推定母数の信頼性にある。本稿では, この問題を2母数対数正規分布を対象として, 月降水量データ数の増加と推定母数の信頼性の関係をエントロピー的立場から把握した。

得られた主な成果を要約すると次のようになる。

1)(1) 式の対数正規分布において, ① λ : 未知, ξ : 既知, ② λ : 既知, ξ : 未知, ③ λ , ξ : 未知, の3つの場合に分け, それぞれ未知とした母数を確率変数とみなして, ベイズ論的立場から求めた事後分布のエントロピーを計算可能な形に導いた。

2) 長野, 松本, 上田の6月, 9月降水量を用いてエントロピーを計算すると, データ数の増加とともに, ①ではエントロピーは単調減少すること, ②, ③ではエントロピーは増加・減少を繰り返しながらも全体的には減少し, データ数が60個程度を超えるとエントロピーの変動が小さくなることがわかった。

今後は, 本稿の手法による他地点の計算例を追加し, ここで得られた成果の一般性を確かめるとともに, 推定母数の信頼性と確率水文学との関連性について議論して, 本研究の土工計画への実用化をはかりたいと考えている。

参考文献

- 1) Benson, M. A. : Uniform Flood-Frequency Estimating Methods for Federal Agencies, Water Resour. Res., Vol.4, No.5, pp.891 ~ 908, 1968.
- 2) 竹内邦良: 不確定性の定式化と分析 ——これからの動向から対処法も含めて——, ②水理・土工計画, 土木学会誌, 第85巻, 第9号, pp.16-20, 1980年。
- 3) 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, 第375号 / II-6, pp.133-141, 1988年。
- 4) Phien, H. N. : A Review of Methods of Parameter Estimation for the Extreme Value Type-1 Distribution, J. Hydrol., pp.251~268, 1987.
- 5) 宝 馨, 高棹琢馬: 水文頻度解析における確率分布モデルの評価規準, 土木学会論文集, 第393号 / II-9, pp.151~160, 1988年。
- 6) 清水 章, 宝 馨, 高棹琢馬: 対数正規分布の母数推定法の比較評価, 土木学会第43回年次学術講演会講演概要集第2部, pp.38 ~ 39, 1988年。
- 7) 宝 馨, 高棹琢馬, 清水 章: 極値分布の母数推定法の比較評価, 土木学会第44回年次学術講演会講演概要集第2部, pp.218~219, 1989年。
- 8) 寒川典昭, 荒木正夫, 上原 剛, 草刈智一: ガンベル分布の推定母数の信頼性評価, 日本気象学会機関誌「天気」(投稿中)。
- 9) 高棹琢馬, 池淵周一: 降水観測がもたらす情報量とその観測網配置計画への応用, 土木学会論文報告集, 第234号, pp.83 ~ 95, 1975年。
- 10) 宮沢光一: 情報・決定理論序説, 岩波書店, pp.126-150, 1971年。

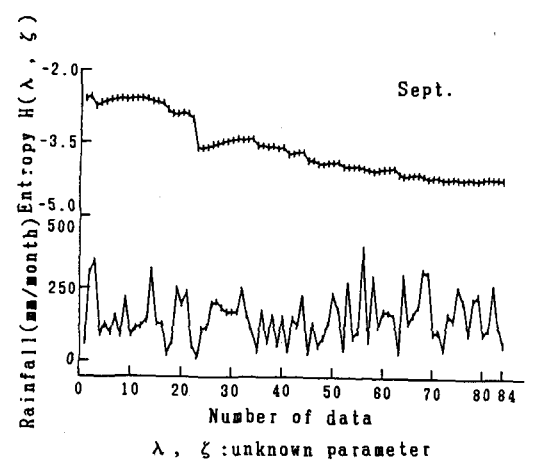
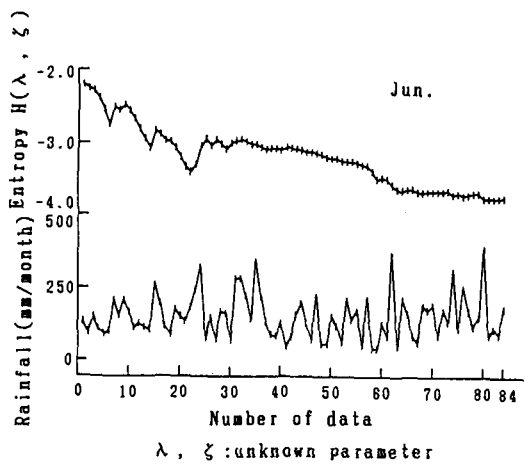
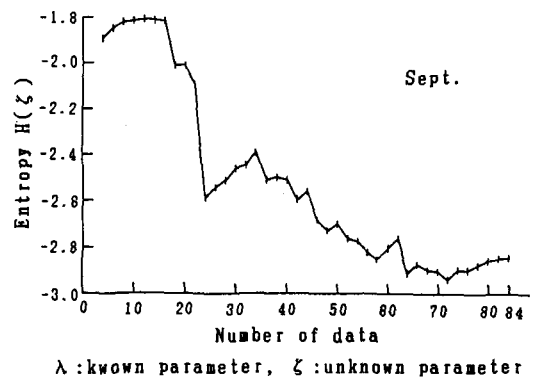
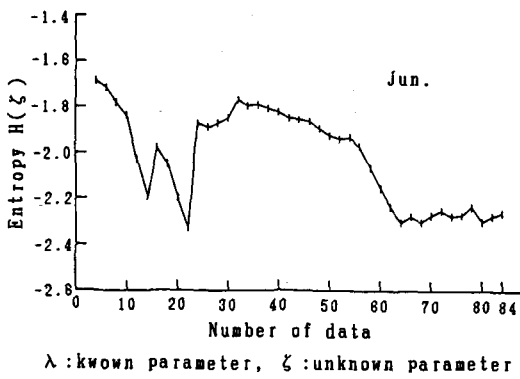
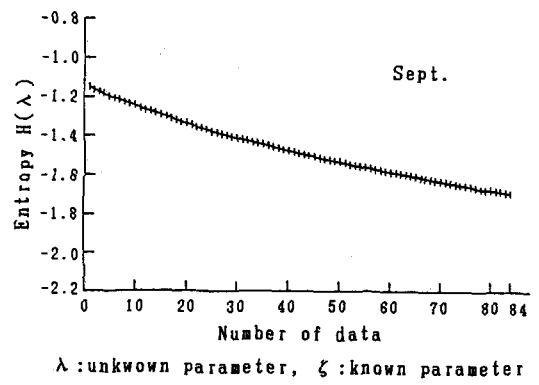
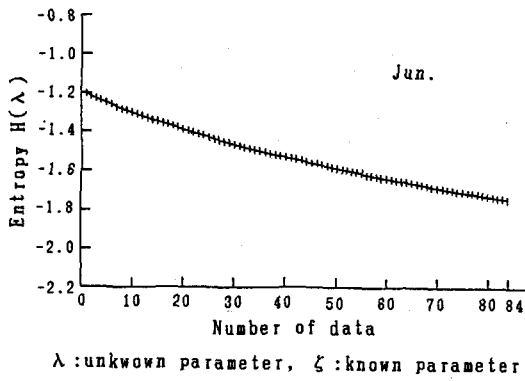


図-1 データ数の増加とエントロピーの関係