

情報と母数の数と最大エントロピー分布

信州大学工学部 正会員 荒木正夫 信州大学工学部 正会員 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 福沢直樹

1. はじめに

最大エントロピー分布を用いて、少ないデータからうまく母集団を推定するためには、より適切な情報を与える必要がある。ここでは、種々の情報の与え方により得られた多数の最大エントロピー分布に、いくつかの評価方法を用いて、如何なる情報を与えた分布が最適かを、極値水量への全体的な適合度および確率水量の安定性の立場から検討する。なお、情報を与える関数は従来の研究¹⁾より、巾関数と指数関数の2種類とし、母数の数は2と3の場合に限定している。

2. 情報の与え方

2母数最大エントロピー分布(MED(2p))は、 $x^s, \exp(-t \cdot x/M)$ (s, t : 正の実数, M : x の平均)の両方の期待値を情報とし、3母数の場合(MED(3p))は、 x^s から2つ、 $\exp(-t \cdot x/M)$ から1つの期待値を情報とした。これらの最大エントロピー分布は次式のようになる。

$$\text{MED}(2p); p(x) = \exp \{-1 - \lambda_0 - \lambda_1 x^s - \lambda_2 \exp(-t \cdot x/M)\} \quad (1)$$

$$\text{MED}(3p); p(x) = \exp \{-1 - \lambda_0 - \lambda_1 x^{s_1} - \lambda_2 x^{s_2} - \lambda_3 \exp(-t \cdot x/M)\} \quad (2)$$

ここで、 s および s_1, s_2 は分布の形状との関係が明らかにされている4次までとし、 t もそれに準じた。

3. 実データへの適用

(1), (2)式の分布を千曲川流域内5地点(長野, 松本, 上田, 飯山, 小海)の年最大1日降水量データに適用した。また、LL(対数尤度), AIC および Jackknife法を用いた評価は文献²⁾を参照した。

a) 長野

(1)式の場合は (s, t) の値をかえたすべての組み合わせについてパラメタ同定を試み、3組を除いた13種類についてパラメタが同定できた。(2)式の場合は (s_1, s_2, t) の値($s_1 = s_2$ は除く)をかえた24種類すべてについてパラメタが同定できた。これら計37種類の分布に、データとの全体的な適合度の客観的な評価規準となるLL, AICを用いて分布に順位付けをした結果、 x^1 と $\exp(-4 \cdot x/M)$ の両方を含んだ情報の組み合わせが上位にランクされた。そこで、良い情報の与え方を見つける第1段階としてLL, AICのそれぞれ上位5位までを選抜した。Table 1にこれらのLL, AICの値を示す。表からわかるようにLLで評価すれば3母数の分布が、AICで評価すれば2母数の分布が上位にくる。しかし、LL, AICの値には大差はなく、分布形状も目視では差異がみられない。したがって、これだけで最良の分布を択一するのは困難である。

そこで次の段階として、水工計画上重要な指標となる確率水量やリターンピリオドの安定性を、2つの手法を用いて調べ最終候補を選定する。はじめにJackknife法を用いて、リターンピリオド $T=30, 50, 100, 200$ 年に対する確率水量の推定誤差、推定値を検討した。Table 1にこの結果を示す。(2, 4)の推定誤差が T の如何に係らず最も小さく、続いて(1, 3), (1, 4, 4)が良い。次にFig. 1はデータを年代の古い順に並べ前半分までのデータの最大値のリターンピリオドについてデータ数の増加に伴う変動をみたものである。(1, 3), (1, 4), (1, 4, 4)等は変動が小さく安定しているが、Jackknife法で最も推定誤差が小さかった(2, 4)はこれらに比べてやや変動が大きく不安定である。結局、Jackknife法を用いた評価、データ数の増加に伴うリターンピリオドの安定性の評価の両方で優れているのは(1, 3), (1, 4, 4)であるといえる。

b) 飯山

飯山の年最大1日降水量を対象とした場合も長野と同様の37種類の分布のパラメタが同定できた。そこでLL, AICを用いて適合度が良い上位5位までに絞り、次いで確率水量、リターンピリオドの安定性を調べた。Table 2にLL, AICの値、Jackknife法による確率水量の推定誤差、推定値を示し、Fig. 2にリターンピ

リオドの変動性を示す。表からわかるように、確率水文学の推定誤差が小さく優れているのは(1,3) (1,4), (1,2,3) である。また、リターンピリオドの変動が相対的に小さいのは(1,4) である。以上のことから、飯山の年最大1日降水量に対しては(1,4) が良い情報を与える組み合わせであることが検証された。

なお、松本、上田の年最大1日降水量には(1,4,4), 小海には(4,4) がそれぞれ優れていた。また現在4母数最大エントロピー分布(MED(4p))を加えた場合についても検討中だが、対象とする地点によっては2, 3母数を越えるものも存在することを付記しておく。

Table 1 LL, AICおよび確率水文学量 (長野)

gr(・)	LL	AIC	Jackknife法によるT年確率水文学量推定値(推定誤差の標準偏差)(mm)			
			30年	50年	100年	200年
1,4	-356.370	716.739	98.57(5.47)	105.76(6.26)	115.35(7.34)	126.53(8.51)
1,3	-356.919	717.838	97.70(5.17)	103.12(5.87)	112.80(6.83)	121.11(7.77)
2,4	-356.978	717.956	97.33(4.87)	101.54(5.39)	109.94(6.03)	117.17(6.63)
1,4,4	-356.333	718.667	98.85(4.99)	107.44(5.85)	116.35(6.77)	123.68(8.27)
1,3,4	-356.358	718.715	98.48(5.09)	106.83(5.87)	113.70(7.25)	124.24(8.94)
1,2,4	-356.369	718.738	99.02(5.23)	107.86(6.14)	115.84(7.67)	125.08(9.69)
1,2,3	-356.845	719.689	97.68(5.26)	104.40(6.04)	113.14(7.10)	122.36(8.21)

Table 2 LL, AICおよび確率水文学量 (飯山)

gr(・)	LL	AIC	Jackknife法によるT年確率水文学量推定値(推定誤差の標準偏差)(mm)			
			30年	50年	100年	200年
1,4	-333.405	670.810	107.83(7.28)	116.57(8.40)	127.04(9.84)	137.72(11.45)
1,2,4	-333.250	672.499	110.15(8.15)	118.47(9.31)	132.52(12.52)	142.84(15.61)
1,3	-334.286	672.572	104.85(7.05)	112.02(8.04)	122.10(9.42)	131.71(10.74)
1,3,4	-333.370	672.740	108.93(7.90)	118.54(9.46)	129.19(11.80)	142.68(14.39)
1,4,4	-333.399	672.798	108.13(7.95)	118.70(9.60)	130.42(12.08)	141.94(14.80)
1,2,3	-334.020	674.040	106.12(7.10)	112.93(8.17)	124.81(9.71)	134.63(11.31)

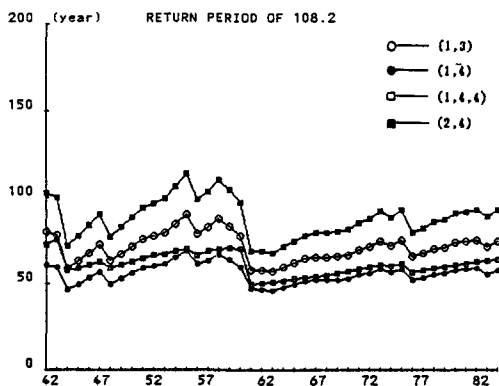


Fig.1 データ数とリターンピリオド (長野)

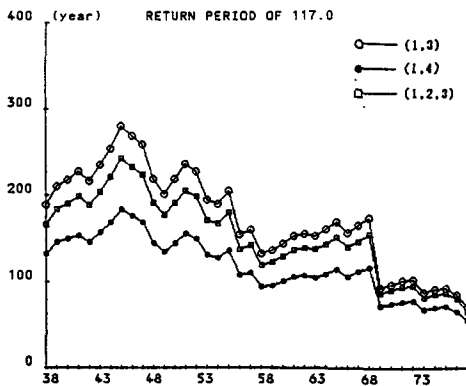


Fig.2 データ数とリターンピリオド (飯山)

4. あとがき

本稿ではいくつかの評価方法を用いて、 $x^{\beta}, \exp(-t \cdot x/M)$ の期待値で情報を与えた2, 3母数の最大エントロピー分布の中から、如何なる分布が最適かを、千曲川流域内5地点の年最大1日降水量データに適用して検証した。今後は同2日, 3日のデータ, および全国の同様のデータに対して検討を行いたい。

- 1) 寒川・荒木・大原：情報の与え方と最大エントロピー分布(その2), 第42回土木学会年講, 第2部.
- 2) 宝・高樺・清水：確率水文学の変動性を規準とした確率分布モデルの評価手順, 第42回土木学会年講, 第2部.