

プロットングポジションの最適化に関する研究

A research on optimization of plotting position

北海学園大学工学部 ○学正員 田守 耕平
北海学園大学工学部 正員 許士 達広

1. まえがき

プロットングポジションは原データの水文資料の順位にいくらの非超過確率値を与えるかを定めるものである。いくつかの式が提案されているが決定的なものが無く、式によって確率水文学の推定値がかなり異なる。例えば100年確率を求めるのに60年間のデータの第1位はワイブルプロットでは1/61、ヘーズンプロットでは1/120の確率になり、比較的水文観測データ数が多いとき確率値付近の差が問題となる。カナンによって整理されて以来40年が経過するが、現在も何が適当であるかは明確に知られてはいない。プロットングポジションは、確率紙を用いた確率水文学の簡易推定法のみでなく、取り扱いが便利なためいろいろな解析に用いられており、極値の確率についても、統計的に回帰分析と対比させて考えられる利点がある。ここではプロットングポジションとして何が適当か、グンベル分布を例に、検討手法について報告する。

2. 既存のプロットングポジション公式

プロットングポジションの主たる研究は1960~70年代になされておりいくつかの式が提案されたが、Cunnane(カナン)によって1)式のように統一的に表現されている。その後も一般化極値分布やピアソン型分布などについての検討は進められているが、基本的な使われ方は変わっていない。

水文学確率値の非超過確率 p はプロットングポジションによって表すことができ次式で示される。

$$p \equiv F(x_i) = \frac{i - \alpha}{n + 1 - 2\alpha} \quad \dots 1)$$

n : 標本数、 x_i : 標本値の小さいものから順に並べた時の i 番目の順序標本値

$F(x_i)$: x_i のプロットングポジション α : 0~1 で表わされる定数。

定数 α として、以下のものがある。

Weibull 公式 : $\alpha = 0$ Hazen 公式 : $\alpha = 0.50$

Blom 公式 : $\alpha = 0.375$ Cunnane 公式 : $\alpha = 0.4$

Gringorten 公式 : $\alpha = 0.44$

上に挙げた公式について、一般的に知られている特徴は以下のとおりである。

Weibull 公式 : 順位標本値の超過確率を与え、 $F(x_i)$ の変化率が一定である。正のひずみ係数を持つ分布形の

分布上端での推定値の偏りが大きいので、使用にあたっては注意する。

Hazen 公式 : 分布形を特定できず、その理論的根拠がしばしば議論されるが、通常考えているよりも適合度は悪くない。

Gringorten 公式 : グンベル分布および指数分布に対し適合度が良い。

Blom 公式 : 正規分布に対して近似度が高い。

Cunnane 公式 : 全ての確率分布形に適用可能な折衷案として提案されたものである。

Gringorten は Weibull、Hazen および Blom の公式が全て1)式の形であることから、1)式の仮定のもとにグンベル分布における標準変量の最大値の理論期待値を用いた超過確率値と1)式のプロットングポジションの最大値が等しいとして $\alpha = 0.44$ を導いている。しかしデータの最大値のみ合わせて適合度が十分であるかという問題が考えられ、分布全体の適合度、確率値の適合度といった観点からも考える必要がある。

3. 本検討のプロットング公式決定の考え方

1) 回帰式との関係

縦軸にデータ x 、横軸に標準変量 ε をとると、直線関係にあるから積率法で係数を求めると

$$a = \sqrt{\frac{S_{xx}}{S_{\varepsilon\varepsilon}}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum (\varepsilon_i - \bar{\varepsilon})^2}} = \frac{\sigma_x}{\sigma_\varepsilon} \quad \dots 2)$$

$$c = \bar{x} - a\bar{\varepsilon} \quad \dots 3)$$

ただし $\bar{x}$: データの平均値 $\bar{\varepsilon}$: 標準変量の平均値
これを用いて積率法の確率値の式は

$$\begin{aligned} x_p &= a\varepsilon_i + c \\ &= \frac{\sigma_x}{\sigma_\varepsilon} \varepsilon_i + \bar{x} - \frac{\sigma_x}{\sigma_\varepsilon} \bar{\varepsilon} \\ &= \bar{x} + \frac{\sigma_x}{\sigma_\varepsilon} (\varepsilon_i - \bar{\varepsilon}) \quad \dots 4) \end{aligned}$$

一方グンベル分布の確率密度関数は以下のように表わされ、非超過確率値は

$$p = F(x) = \exp \left[- \exp \left(- \frac{x - c}{a} \right) \right] \quad \dots 5)$$

x : 水文学、 a : 尺度母数、 c : 位置母数

水文学 x の確率値 x_p は標準変量 ε_p の一次式であり、

確率値は次式で表される

$$x_p = a\varepsilon_p + c = x + \frac{\sqrt{6}}{\pi} \sigma_x (\varepsilon_p - 0.5772) \quad \dots 6)$$

$$\varepsilon_p = -\ln[-\ln(p)]$$

4)式と比較すると

$$\bar{\varepsilon} = 0.5772 \quad \dots 7) \quad \sigma_\varepsilon = \frac{\pi}{\sqrt{6}} = 1.2825 \quad \dots 8)$$

グンベル分布の標準変量の条件がプロットによる直線回帰に対する条件であることが分かる。

2)プロットング公式の形状

プロットング公式は従来1)式が用いられているが、ここでは一般性を持つ以下の形とする。

$$p = \frac{i - \alpha}{N + 1 - \beta} \quad \dots 9)$$

この式の α 、 β を定めることになるが、 α はプロットの中心の位置に関係し、特に端のプロットに効いてくる。 β は尺度に関係し、主にプロットの幅に影響する。またデータ数 N により α 、 β が変わることが予想される。グンベル分布では標準変量は

$$\varepsilon_i = -\ln[-\ln(p_i)] = -\ln\left[-\ln\left(\frac{i - \alpha}{N + 1 - \beta}\right)\right] \quad 10)$$

であり、その平均値と分散が7)、8)式を満たすことが、プロットングポジションの α 、 β を決めていく条件となる。これは積率法による確率式による理論解とプロットングポジションからの積率法の回帰式を一致させていることになる。

設定したデータ数 N ごとに10)式で α 、 β の値を変えて、確率分布の標準変量 ε_i を算出し、標準変量の平均値 $\bar{\varepsilon}$ 標準変量の分散 σ_ε が7)8)式に示した固有の値に最も近づくように、トライアルして α 、 β の最適な組み合わせを定める。過去のプロットングポジションの式では $\alpha = 0 \sim 0.5$ 、 $\beta = 0 \sim 1.0$ であるが、非超過確率 p あるいは超過確率 $1 - p$ は対数がとられ、対数の真数が正である必要がある。これからトライアルにあたっての実際の条件は $0 < \alpha < 1$ 、および $0 < \beta < 1 + \alpha$ となり、より大きい範囲が対象となる。

4. 計算法

σ_ε と $\bar{\varepsilon}$ の両方の条件7)、8)式を同時に満足する α 、 β の組み合わせは存在しないので、両方を考えた評価指標が必要である。ここでは次の4つの方法で α 、 β をトライアルして求めている。

1)平均と分散の理論値を用いて

$$\sigma_\varepsilon^2 = \frac{1}{N} \sum_{i=1}^N (\varepsilon_i - \mu_\varepsilon)^2 = \frac{1}{N} \left[\sum_{i=1}^N \varepsilon_i^2 - N \mu_\varepsilon^2 \right] \quad 11)$$

であるから

$$\frac{1}{N} \sum_{i=1}^N \varepsilon_i^2 = \sigma_\varepsilon^2 + \mu_\varepsilon^2 = 1.6449 + 0.5772^2 = 1.9781 \quad \dots 12)$$

を満たす α 、 β の組み合わせを求める。

2)同様に $\frac{1}{N} \sum_{i=1}^N \varepsilon_i$ を計算して

$$s_\varepsilon^2 = \frac{1}{N-1} \sum_{i=1}^N (\varepsilon_i - \bar{\varepsilon})^2 = \sigma_\varepsilon^2 = 1.6449 \quad \dots 13)$$

よりトライアルで求める。

3)確率式4)式から正確な確率値を求めるプロットは

$$x_p = \bar{x} + \sigma_x K_p \quad K_p = \frac{\varepsilon_p - \bar{\varepsilon}}{s_\varepsilon} \quad \dots 14)$$

となるとき K_p が正確なときである。

たとえば1/50確率値のときに、非超過確率 $p = 0.98$

$$\varepsilon_{1/50} = -\ln(-\ln 0.98) = 3.9019$$

このとき K の理論値は

$$K_{1/50} = \frac{3.9019 - 0.5772}{1.2825} = 2.5924$$

これに対し K の計算値 K^* は

$$K_{1/50}^* = \frac{3.9019 - \bar{\varepsilon}}{s_\varepsilon}$$

であり α 、 β を仮定して $K_p = K_p^*$ となる値を求める。

1つの確率でなく、使用する範囲の確率 p で求めるとき

$$TK = \sum_p (K_p - K_p^*)^2 \quad \dots 15)$$

が最小になる α 、 β とする。

4)上記の1)や2)の方法では求められる α 、 β では σ_ε^2 に比べ $\bar{\varepsilon}$ の誤差が大きくなる傾向が見られる。

このため、 σ_ε^2 と $\bar{\varepsilon}$ の誤差が同程度になるように

$$E = \left(\frac{s_\varepsilon - 1.2825}{1.2825} \right)^2 + \left(\frac{\bar{\varepsilon} - 0.5772}{0.5772} \right)^2 \quad \dots 16)$$

が最小になるように α 、 β を定める。

計算結果については現在検証中であるが、計算方法によって最適の α 、 β の違いがあり、いずれも β は1を超え、 α もカナン式より大きな値となっている。 β を1までに抑えるとGringortenの式にやや近づく。

参考文献

- 1) C. Cunnane: Unbiased Plotting Positions - A Review, Journal of Hydrology, 37 (1978) 205-222
- 2) Irving I. Gringorten: A Plotting Rule for Extreme Probability Paper, Journal of Geophysical Research vol. 68, no. 3