

プロットイングポジションを用いた水文確率の算定

A research on statistic hydrologic using plotting position

○北海学園大学社会環境工学科 中村皓平

北海学園大学社会環境工学科 佐藤大介

北海学園大学社会環境工学科 フェロー 許士達広 (Tatsuhiko Kyosh

i)

1. はじめに

水文確率値の算定については、モデル分布形状をどのようにするかとともに、その母数算定手法が影響するため過去に多くの研究がされてきた。比較的計算しやすい積率法が発達し、高次モーメントの誤差解消のためにL積率法が考案された。最尤法は数学的には最も正しい方法であるが、方程式が複雑でニュートン法などによる計算が必要となり、収束しない場合もあるため、実用的にはあまり普及していない。最尤法で誤差分布が正規分布（左右対称）であれば最小二乗法が等価で数学的に最適な方法となる。実際モデルの適合性評価に用いられる決定係数や AIC、SLSC は最小二乗誤差を用いたものであり、プロットイングポジションの定数が統一されていれば評価指標に対し最小二乗法が最も適した手法になる。最小二乗法は単なる直線式であり計算がきわめて単純であるが、問題点はプロットイングポジションの式である。プロットイングポジションは計算機が未発達の時代に確率紙にプロットして使われたもので、時代遅れの印象を持つ人もいるであろうが、最小二乗法でデータの生起確率を出すときには欠かせないものである。ここではプロットイングポジションの定数 α の算定とそれを用いた最小二乗法の精度について検討する。

2. 既往研究

プロットイングポジションは原データの水文資料の順位に非超過確率値を与える式を定めるものである。カナンによって整理されて以来 40 年以上が経過し、その後も検証されているが決定的なものが無く、いくつかのプロットが並存して使われている。一般的には以下の式である。式の p_i は x_i の点の非超過確率を示す。

$$p_i = F(x_i) = \frac{i - \alpha}{N + 1 - 2\alpha} \quad \dots (1)$$

N : 標本数 x_i : 標本を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 のプロットイングポジション、

α : 0~1 で表わされる定数。

確率値は例えば 2 母数であれば

$$x_p = a\varepsilon(p) + c \quad (2)$$

(1) 式を(2)式の標準変数 ε の中の非超過確率 p として用いることにより、確率値は p の関数として最小自乗法で求められる。

プロットイングポジションの研究は Hazen 式に始まり、

特に 1940~70 年代に多く行われている。数多くの公式が提案され、公式により確率値に差が生じるため、Cunnane(1977)によって前節の(1)式のように統一的に表現された。(1) 式における定数 α として 主なものとその一般的に知られている特徴は以下のようになる。

① Hazen 公式 (1914 年) : $\alpha = 0.50$

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。

② Weibull 公式 (1939 年) : $\alpha = 0$

順位標本値の超過確率を与え、 $F(x_i)$ の変化率が一樣である。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を要する。

③ Blom 公式 (1958 年) : $\alpha = 3/8$

正規分布に対して近似度が高いとされている。

④ Gringorten 公式 (1963 年) : $\alpha = 0.44$

グンベル分布および指数分布に対し適合度が良いとされている。

⑤ Cunnane 公式 (1977) : $\alpha = 0.4$

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。その後の検証でも比較的適合度が良いことから Hazen プロットと共に標準的な式として利用されている。このほかに、重みつき積率の簡易計算法として考えられた Landwehr の方法があり、これは分子の α が 0.35、分母の α が 0.5 である。

Cunnane 以降 40 年が経過し、その間 3 母数の一般化極値分布や対数ピアソンⅢ型分布に対するものなどいくつかの論文が発表されているが、基本的な見解はあまり変化していない。

3. 本研究の方法

1) 基本的考え方

確率値の真値は分からないためデータにあった分布形状を探すのであるが、それぞれのモデル分布において算定法ごとに母数の理論値が存在する。それらは非超過確率 p の関数として表すことができる。 p に(1)式のプロットイングポジションを用いて理論値に合わせれば、その分布の母数に合った α が算定される。具体的には α を仮定して p から標準変量を算出し、それから得られる平均値と標準偏差、あるいは歪み度と尖り度を算出する。 α を変化させて繰り返し計算により、計算した平均値や標準偏差が理論値に一致するように α を定める。しかし最適な α は平均と標準偏差など母数毎に異なるため、それぞれの誤差を合成したものが最小となるように定める。

合成の方法は確率値 x_p の誤差を最小にする方法とし、
確率値の誤差伝搬として以下の3種類が考えられる。

誤差伝搬 1

$$dX_p^2 = \left(\frac{\partial K_p}{\partial c}\right)^2 dc^2 + \left(\frac{\partial K_p}{\partial a}\right)^2 da^2$$

誤差伝搬 2

$$dX_p = \left(\frac{\partial X_p}{\partial c}\right) dc + \left(\frac{\partial X_p}{\partial a}\right) da$$

誤差伝搬 3

$$dX_p^2 = \left(\frac{\partial K_p}{\partial c}\right)^2 dc^2 + \left(\frac{\partial K_p}{\partial a}\right)^2 da^2 + 2\left(\frac{\partial K_p}{\partial c}\right)\left(\frac{\partial K_p}{\partial a}\right) dadc$$

2) 各分布および算定法

① グンベル分布・積率法 (2母数のみ)

$$\begin{aligned} x_p &= c + a\varepsilon_p = \bar{x} + \frac{s_x}{s_\varepsilon}(\varepsilon_p - \bar{\varepsilon}) \\ &= \bar{x} + s_x \frac{\varepsilon_p - \bar{\varepsilon}}{s_\varepsilon} \\ K_p &= \frac{\varepsilon_p - \bar{\varepsilon}}{s_\varepsilon} \quad \varepsilon_i = -\ln[-\ln(p_i)] \\ \frac{\partial K_p}{\partial \bar{\varepsilon}} &= -\frac{1}{s_\varepsilon} \quad \frac{\partial K_p}{\partial s_\varepsilon} = -\frac{(\varepsilon_p - \bar{\varepsilon})}{s_\varepsilon^2} \\ dK_p &= \left(\frac{\partial K_p}{\partial \bar{\varepsilon}}\right) d\bar{\varepsilon} + \left(\frac{\partial K_p}{\partial s_\varepsilon}\right) ds_\varepsilon \end{aligned}$$

$$\begin{aligned} d\bar{\varepsilon} &= \bar{\varepsilon} - \mu_\varepsilon \quad ds_\varepsilon = s_\varepsilon - \sigma_\varepsilon \\ \sigma_\varepsilon &= \frac{\pi}{\sqrt{6}} \quad \mu_\varepsilon = \gamma = 0.577216 \end{aligned}$$

② 正規分布・積率法

$$x_p = \bar{x} + \frac{\sigma_x}{\sigma_\varepsilon}(\varepsilon_p - \mu_\varepsilon) = \bar{x} + \sigma_x \varepsilon_p$$

$$dK_p = \left(\frac{\partial K_p}{\partial \bar{\varepsilon}}\right) d\bar{\varepsilon} + \left(\frac{\partial K_p}{\partial s_\varepsilon}\right) ds_\varepsilon$$

$$= \frac{1}{s_\varepsilon} \bar{\varepsilon} + \frac{(\varepsilon_p - \bar{\varepsilon})}{s_\varepsilon^2} (s_\varepsilon - 1)$$

$$\mu_\varepsilon = 0, \sigma_\varepsilon = 1$$

$$d\bar{\varepsilon} = \bar{\varepsilon} - 0 \quad ds_\varepsilon = s_\varepsilon - 1$$

③ グンベル分布・積率法 (4母数)

グンベル分布自体は2母数であるが、ひずみ度や尖り度も当然存在する。それに対しても α が合うように誤差伝播を合成する。

$$dK_p = \frac{\partial K_p}{\partial \varepsilon} d\varepsilon + \frac{\partial K_p}{\partial s_\varepsilon} ds_\varepsilon + \frac{\partial K_p}{\partial \gamma} d\gamma + \frac{\partial K_p}{\partial \beta} d\beta$$

$$\gamma_\varepsilon = M_3/S^3 \quad d\gamma = \gamma_\varepsilon - 1.1396$$

$$\beta_\varepsilon = \frac{M_4}{S^4} - 3 \quad d\beta = \beta_\varepsilon - \frac{12}{5}$$

4. 計算結果

各分布の平均、標準偏差について合成した誤差伝搬を最小にするプロットング定数 α を示す。データ数Nごとに値は変化する。

① 正規分布

	α		
データ数N	誤差伝播1	誤差伝播2	誤差伝播3
20	0.522	0.522	0.522
50	0.518	0.518	0.518
100	0.516	0.516	0.516
200	0.514	0.514	0.514
1000	0.507	0.507	0.507

② グンベル分布

	α		
データ数N	誤差伝播1	誤差伝播2	誤差伝播3
20	0.586	0.588	0.588
50	0.578	0.579	0.579
100	0.573	0.574	0.574
200	0.569	0.569	0.569
1000	0.560	0.560	0.560

③ 対数正規分布

	α		
データ数N	誤差伝播1	誤差伝播2	誤差伝播3
20	0.763	0.757	0.757
50	0.741	0.737	0.737
100	0.728	0.724	0.724
200	0.716	0.712	0.712
1000	0.693	0.691	0.691

④ 一般化極値分布

k = -0.2 歪み度 3.5350

k = -0.1 歪み度 1.9103

	α	
データ数N	平均	分散
30	0.642	0.712
50	0.631	0.706
100	0.620	0.699
200	0.612	0.693
1000	0.598	0.682

	α	
データ数N	平均	分散
30	0.622	0.644
50	0.610	0.639
100	0.598	0.632
200	0.589	0.627
1000	0.574	0.616

k = 0.1 歪み度 0.6376

k = 0.2 歪み度 0.2541

	α	
データ数N	平均	分散
30	0.586	0.532
50	0.574	0.529
100	0.562	0.525
200	0.552	0.521
1000	0.537	0.514

	α	
データ数N	平均	分散
30	0.553	0.500
50	0.540	0.497
100	0.526	0.493
200	0.515	0.491
1000	0.496	0.486

一般的にプロットングポジションの定数 α は0.5以下と考えられているが、いずれも0.5より大きな値となる。一般化極値分布の場合は平均、分散(標準偏差)のみ算出しているが、ひずみ度毎に分布形状が変わるので α が変化することが分かる。積率法ばかりでなくL積率法や最尤法でも算定でき、異なる α が算定される。