水文確率の適合度と安定性に関する一考察

A study on the stability and adaptability of hydrological probability distribution model

北海学園大学工学部 学生員 中矢 拓磨 (Takuma Nakaya)

北海学園大学工学部 学生員 横井 潤 (Jun Yokoi)

北海学園大学工学部 正員 許士 達広 (Tatsuhiko Kyoshi)

1. はじめに

河川計画に用いられる水文確率値を求める上で、多くの確率分布モデルや母数推定法が利用されている。これらのうち何が最適であるか判定するために、主として原データの分布関数への適合度が用いられる。また最近では計画策定上標本数の違いにより確率値が大きく変動しないようなもの(変動性最小)が望ましいという考え方が出されている。前者については分布の²検定やSLSC、AICといったものが用いられ、後者にはジャックナイフ法やブートストラップ法といったリサンプリング法による誤差推定分散が指標とされている。現在どの分布モデルが望ましいかという問題について明確な答えはなく、実用的にはデータ毎に各種のモデルを計算し、値の近いものの中から選定しているのが実態である。本研究ではこれらを踏まえ、分布の変動性について確率分布の構造から考察を行なうものである。

2. 検討対象モデル

水流量に用いられる確率分布には多くのモデルが考案されているが、通常用いられる年最大日雨量などの毎年確率分布は、対数正規分布、極値分布、ガンマ分布に大別される。ここではそれぞれの中で代表的な以下の分布を対象とする。

・3母数対数正規分布

$$f(x) = \frac{1}{(x-a)\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\ln(x-a)-\mu_y}{\sigma_y}\right)^2\right]$$

$$y = \mu_y + \sigma_y \cdot Z_p \quad (1)$$

ただし $y = \ln(x-a)$

a : 分布下限値 μ_y : y の平均 σ_y : y の分散

$$Z_p = -\sqrt{\frac{t^2[(4t+100)t+205]}{[(2t+56)t+192]t+131}} \quad (2)$$

$$t = -\ln(2p) \quad p: \text{非超過確率}$$

・グンベル分布

$$f(x) = \frac{1}{a} \exp\left[-\frac{x-c}{a} - \exp\left(-\frac{x-c}{a}\right)\right]$$

$$x = c + a \cdot G_p \quad (3)$$

ただし $G_p = -\ln[-\ln(p)]$ p : 非超過確率

$$\mu_x = c + 0.5772a \quad \sigma_x = \pi a / \sqrt{6}$$

または $x = \mu_x + \sigma_x \cdot K$ (4)

$$\text{ただし } K = \sqrt{6} / \pi (G_p - 0.5772)$$

・対数ピアソン 型分布

$$f(x) = \frac{1}{a|\Gamma(b)x|} \left(\frac{\ln x - c}{a}\right)^{b-1} \exp\left(-\frac{\ln x - c}{a}\right)$$

$$y = \mu_y + \sigma_y \cdot K_p \quad (5)$$

ただし $y = \ln x$

$$K_p = \frac{2}{\gamma_y} \left(1 + \frac{\gamma_y Z_p}{6} - \frac{\gamma_y^2}{36}\right)^3 - \frac{2}{\gamma_y} \quad (6)$$

p : 非超過確率 γ_y : ひずみ係数 Z_p : 標準正規変量

用語の使い方には多少課題があるが Z_p 、 G_p 、 K_p を標準変量と統一して呼ぶことにすると、主要な分布は非超過確率に対応する Z_p などの標準変量と水流量の直線式で表され、直線回帰の問題と対比して考えることが出来る。

3. 回帰式での表現

1変数の直線の回帰式は平均値を通る直線として

$$y = ax + b = \bar{y} + a(x - \bar{x}) \quad (a, b: \text{定数}) \quad 7)$$

のように表される。この直線の信頼区間は

$$\hat{y}_0 = \bar{y} \pm t(n-2, \alpha) \text{Se} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \quad 8)$$

また新たな予測値の信頼区間は

$$\hat{y}_0 = \bar{y} \pm t(n-2, \alpha) \text{Se} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \quad 9)$$

で表せる。ただし

$t(n-2, \alpha)$: t 分布の超過確率 $\alpha/2$ に対する標準変量 n : 標本数

$$\text{Se} = \sqrt{\frac{\sum(\hat{y}_i - y_i)^2}{n-2}} \quad S_{xx} = \sum(x_i - \bar{x})^2$$

式における Se は残差平方和に関係し、直線の適合度を示す。

一方 $\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$ の中の $\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}$ はテコ比といわれ、その測定値が回帰直線に与える影響を表す。テコ比は予測値 $\hat{y}_0$ における変数 x_0 が、 x の平均値 $\bar{x}$ からどの

くらい離れているかなどにより変化し、直線の適合度とは別に予測値の信頼区間に影響する。

もしこれを水文確率の式に適用すれば、回帰式 9) の y は水文観測値 x 、回帰式 9) の x は標準変量の Z_p や G_p 、 K_p となる。

水文学の各測定値毎の非超過確率 p はプロットングポジションで表すことが出来る。プロットングポジションは次式で示される。

$$p = F(x_i) = \frac{i - \alpha}{n + 1 - 2\alpha} \quad 10)$$

n : 標本数 x_i : i 番目の順位標本値

は 0~1 の定数で公式ではワイブルが $\alpha = 0$

ヘーズン $\alpha = 1/2$ 、カナン $\alpha = 2/5$ などとなる。

式 1) ~ 6) に示すように Z_p 、 G_p 、 K_p は p により定まることから、これらは標本数 n の関数である。従って式 8) および 9) のテコ比の部分は、確率分布形毎に標本数と求める確率に影響される固有の値をとり、対数正規分布やゲンベル分布では測定値に関係しない。したがって式 9) に示されるように水文確率値の信頼性は、モデルの分布形に固有の確率値の信頼性を示す値

$$h_0 = \sqrt{1 + \frac{1}{n} + \frac{(z_0 - \bar{z})^2}{S_{zz}}} \quad \text{と観測データごとに定まる分布}$$

の適合度を表す誤差の標準偏差

$$\text{Se} = \sqrt{\frac{\sum(\hat{x}_i - x_i)^2}{n-2}} \quad \text{の積で表され、大きいものが信頼}$$

性が低い。

4. 分布毎の特性

前述した 3 つの確率分布について前節の h_0 を計算する。プロットはカナン式を用いたものを表 - 1 に示す。

表 - 1 h_0 の計算

求める確率	1 / 1 0 0		1 / 4 0 0	
データ数	2 0	1 0 0	2 0	1 0 0
対数正規	1.1633	1.0323	1.2213	1.0445
ゲンベル	1.0939	1.0177	1.1167	1.0221
ピアソン	1.1412	1.0422	1.1828	1.0640

示した計算数が少ないので判りにくいですが、求める確率が大きいほうが分布の平均値との差が大きく、分布による違いが出やすい。またデータ数が多くなるとテコ比が 1 に比べて小さくなり、差が生じない。この例では h_0 が最も小さいのがゲンベルであり、対数正規とピアソンはデータ数により大小の順位が変わる。これが確率値の安定性を示す一つの指標と考えられる。

信頼限界の誤差分散 $V(y)$ 等について、旭川の既往年最大日雨量を用いて計算した例を表 - 2 に示す。モデルどうしの比較を行なう場合、データの大きさの違いや、実数をとったモデルと対数を取ったモデルとの比較のために標準化が必要である。このため確率値の誤差分散を、99% と 1% の水文確率値の差で割ったものが比較には有効と考えられ、これは一般に用いられている SLSC と同等の意味を持つ。

$$DV(y) = \frac{\sqrt{\frac{1}{n} \sum(\hat{y}_i - y_i)^2}}{|y_{0.99} - y_{0.01}|} \quad 11)$$

y : データの値、対数の場合は対数値でよい。

$y_{0.99}$ 、 $y_{0.01}$: それぞれ非超過確率 0.99 と 0.01 の確率値、ただし対数の場合は対数値でよい。

表 - 2 確率値の分散 $V(y)$ (旭川 1/100)

モデル	対数正規		ゲンベル		ピアソン	
データ年数	2 0	1 0 0	2 0	1 0 0	2 0	1 0 0
h_0	1.1633	1.0445	1.0939	1.0177	1.1412	1.0422
Se	1.9533	3.6938	1.8509	6.1783	1.8583	3.8533
$V(y)$	2.2723	3.8581	2.0247	6.2877	2.1207	4.0160
$DV(y)$	0.0302	0.0246	0.0269	0.0436	0.0274	0.0241

5. おわりに

水文確率分布について、回帰式と対比させて説明を試みた。 h_0 や $DV(y)$ により確率分布の特性がある程度説明できると思われ、さらに実際のデータで検証する必要がある。

参考文献

- ・星清：水文統計解析 開発土木研究所月報 NO.540(p15-47)
- ・竹内啓監修：SAS による回帰解析(p193-203)
- ・石黒正義、他：応用水文統計学(p64-96)