

プロットイングポジションを用いた非毎年確率の研究

Study of suitable quantiles of probability distribution for partial duration series using plotting position

北海学園大学社会環境工学科 フェロー 許士達広 (Tatsuhiko Kyoshi)
 北海学園大学社会環境工学科 学生員 川森亮輔 (Ryousuke Kawamori)
 北海学園大学社会環境工学科 学生員 三上沙希 (Saki Mikami)

1. まえがき

従来河川の計画の分野では、雨量や流量の年最大値を用いる毎年確率が用いられてきたが、近年、河川整備計画検討において、降雨の毎年確率の適合度が低い場合の検討事項として、非毎年確率が位置付けられ非毎年確率の使用頻度が高まっている。また対象とする規模の現象が必ずしも毎年発生しない場合など、毎年確率値が当てはめにくい時には非毎年確率値で考える必要がある。非毎年確率は、生じた年に関係なく、閾値と呼ばれる一定値以上の水文量を用いた計算法である。閾値は非毎年確率の確率値算定に大きく影響するが、設定方法に適切なものがないことが、非毎年確率の欠点として極値統計学において、長年の大きな課題となっている。本研究では非毎年確率の課題である閾値の設定方法について回帰直線と棄却検定の考え方を用いて検討する。

2. プロットイングポジションを用いた非毎年確率値

非毎年確率値の計算法には非毎年データの非超過確率の分布として、一般的に指数分布と一般化パレート分布が用いられている。水文学においては、これらの分布を毎年確率と同様の確率年(再現期間)を持った水文量として表すために、ポアソン分布で変換し、それぞれグンベル分布、および一般化極値分布の形にして用いている。これらの分布の定数は一般的に重み付きのL積率法などで求められるが、ここでは閾値の算定のためにプロットイングポジションを用いた最小二乗法による算定する方法を用いた。

1) プロットイングポジションを用いた回帰式

非毎年確率を回帰直線として考え、確率分布の標準変量をプロットイングポジションで表す。非毎年の水文データを縦軸、分布の標準変量を横軸に取りプロットする。

非毎年確率を毎年に換算した下から*i*番目の非超過確率*p_i*は

$$p_i = \exp\{-\lambda[1 - G(x)]\} = \exp\left\{-\frac{M}{N}\left(1 - \frac{i - \alpha}{M + 1 - 2\alpha}\right)\right\} \quad \dots 1)$$

M: 閾値以上のデータ個数

N: データ年数 *i*: プロットイングポジション

α : プロットイングポジションの定数

G(x): 非超過確率の分布関数, cdf

λ : ポアソン分布の閾値以上の値の年平均生起回数
 指数分布によるグンベル分布の確率値は

$$x_p = c + a\{-\ln(-\ln(p))\} = c + a\left[-\ln\left\{\frac{M}{N}\left(1 - \frac{i - \alpha}{M + 1 - 2\alpha}\right)\right\}\right] \quad \dots 2)$$

一般化パレート分布による一般化極値分布の確率値は

$$x_p = c + a\left[\frac{1}{k}\left\{1 - \{-\ln(p)\}^k\right\}\right] = c + a\left[\frac{1}{k}\left\{1 - \left\{\frac{M}{N}\left(1 - \frac{i - \alpha}{M + 1 - 2\alpha}\right)\right\}^k\right\}\right] \quad \dots 3)$$

いずれも [] 内が標準変数 ε_i であり、以下の直線関係となる。

$$x_p = a * \varepsilon_i + c \quad \dots 4)$$

2) 定数決定法

一般化極値分布の*k*値はデータから歪み係数合わせで先に求める。歪み係数合わせとは4)式が成り立つときデータと標準変量のひずみ係数が等しくなることを利用するもので、以下の式で示され、トライアルで*k*値が求められる。

$$\varepsilon_i = \frac{1}{k}\left[1 - \left\{\frac{M}{N}\left(1 - \frac{i - \alpha}{M + 1 - 2\alpha}\right)\right\}^k\right] \quad \dots 5)$$

$$\frac{\sum_{i=1}^M (x_i - \bar{x})^3}{\left\{\sum_{i=1}^M (x_i - \bar{x})^2\right\}^{3/2}} = \frac{\sum_{i=1}^M (\varepsilon_i - \bar{\varepsilon})^3}{\left\{\sum_{i=1}^M (\varepsilon_i - \bar{\varepsilon})^2\right\}^{3/2}} \quad \dots 6)$$

ε_i : 一般化極値分布の標準変数 x_i : 下から*i*番目の水文

データ値 M: 閾値以上の対象データ数

回帰式による確率分布の母数推定法は一般の直線回帰同様に以下の式となる。

$$a = \frac{S_{x\varepsilon}}{S_{\varepsilon\varepsilon}} = \frac{\sum (x_i - \bar{x}) \sum (\varepsilon_i - \bar{\varepsilon})}{\sum (\varepsilon_i - \bar{\varepsilon})^2} \quad \dots 7)$$

$$c = \bar{x} - a\bar{\varepsilon} \quad \dots 8)$$

$\bar{x}$: データの平均値 $\bar{\varepsilon}$: 標準変数の平均値

3. 閾値の決定

1) 決定の考え方

閾値は確率値の推定の観点から、ある値から上のデータと下のデータが別な確率値を与える境界と考えられる。確率値の推定のために、標準変量を横軸、水文データを縦軸としてプロットした場合、当てはめた回帰直線のどの値から上が同じ確率値を与えるか、下限の値が閾値 x_0 である。分布が等しい範囲はお

折点の最適化のために情報量基準 AIC を使い、AIC が極小となる折れ点、あるいは大きく変化する折れ点の値を求める。これらの折れ点は閾値の候補点であると考えられる。2 本の直線に