

極値付近の適合度を重視した確率分布の検討

The research on the characteristic distribution that a level of adaptation around the extreme value is improved

北海学園大学工学部 ○学生員 三本木公士(Hitoshi Sambongi)
北海学園大学工学部 正員 許士 達広(Tatsuhiko kyosi)

1. まえがき

河川計画のための高水流量の決定には、主として降雨の超過確率値から時間分布により流量に変換したものが用いられている。降雨の超過確率値は、一般的に毎年確率で求められ、何十年かの既往観測資料から日雨量や3日雨量の年最大値を抽出し、その正規確率分布から外挿により100年あるいは150年といった確率値を推定している。確率分布には対数正規分布や極値分布、およびその母数の推定法には岩井法や積率法、L積率法などの種々のものが考えられており、その適合度評価のためにSLSCなどの評価法が提案されている。

超過確率値は、確率分布における端部である袖の部分の推定しているため、この非常に低い確率の部分の観測資料との適合度が確率値の信頼性を決定することになる。一方、従来確率分布や母数推定はそれぞれの観測標本を同等に扱い、それらと適合することにより決定・評価されている。確率水文学は用いた標本数、確率分布形及び母数推定法に大きく変動することが知られており、これが計画上の問題となっている。一定数以上の標本数が得られた場合、求める確率値付近のデータとの適合度を重視する手法を用いることにより手法の違いによる変動が減少することが考えられる。ここでは、そのための概略検討を示す。

2. 確率値付近のデータ重視の考え方

近年データの蓄積が増加するにつれて、求める確率値に近い年数のデータが使えるケースが出ている。

例えば、100年間のデータがあり100年確率値を論ずるときに一番大きな値によりおおよその推定が出来るが、一番小さな値からは不可能である。同様に100年間のデータがあり、150年確率値(極大値)を求めるとき、100年間の大きい方の50年間のデータから得られる情報は、100年間の小さい方の50年間のデータから得られる情報よりはるかに大きい。これは情報量理論で確率値 p であるための情報量は $-1\log_2 p$ であり、 $1/100$ の確率の情報量は $1/2$ の確率の情報量の6倍であるといったことにも通じると思われ、1年や2年に1回生ずる雨より、50年や100年に1回の雨の情報を重視することが精度向上につながると考えられる。

3. 最小二乗法の適用

データによって適合に関する重みを変えた確率値を求める検討の一環として、データの一部から確率値を求める方法を考えてみる。最尤法や積率法からこういった扱いは難しく、例えば大きいほうの値だけで積率法で母

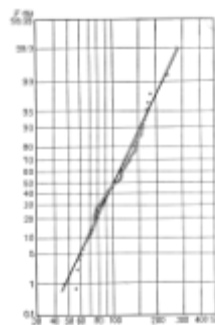


図-1 対数正規確率紙

数を求めれば、平均値の大きく違う分布を考えることになる。これを避けるためここでは最小二乗法を用いる。図-1の対数正規確率紙において縦軸はプロットングポジション、横軸は実測データとすると、対数正規確率の最小二乗法は、この図に直線を当てはめることと同じである。

プロットングポジション公式は以下のように与えられる。ここでは、ワイブル・ハーゼン・カナンの3種類を用いる。

$$p = \frac{i - \alpha}{N + 1 - 2\alpha}$$

ワイブル： $\alpha=0$

ハーゼン： $\alpha=1/2$

カナン： $\alpha=2/5$

3母数正規分布(LN3)は以下の式で表される。
確率レベル p に対するクオンタイル(確率値)

$$x_p = a + \exp(\mu_y + \sigma_y \cdot z_p)$$

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

$$y = -\ln(2p) \quad (10^{-7} < p < 0.5)$$

3母数で最小二乗法を扱うと複雑になるので、母数 a は岩井法の値を使い、2母数で計算をする。

$$S = \sum [\ln(x_i - a) - (\mu_y + \sigma_y \cdot z_i)]^2$$

これを μ_y と σ_y で偏微分を行い、その結果を0として得られる2式を展開し母数を求める。

グンベル分布は以下の式で表される。

確率レベル p に対するクオンタイル(確率値)

$$x_p = c - a \ln[-\ln(p)]$$

$y = -\ln(-\ln p)$ として、ここではプロットングポジションと観測値の間の最小二乗法を考える。

$$S = \sum [x_i - (c + a \cdot y_i)]^2$$

これを c と a で偏微分をし、その結果を0として得られる2式を展開し母数を求める。

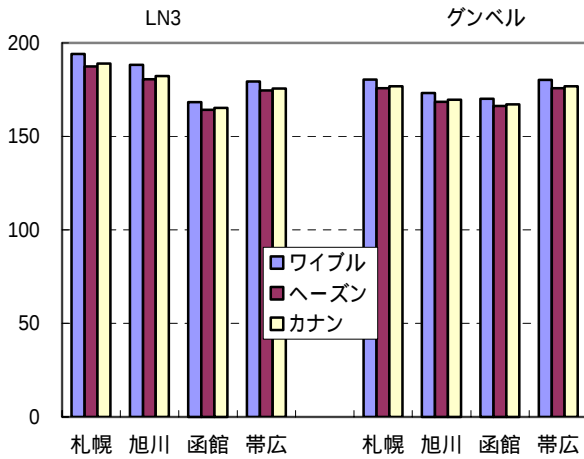
SLSCとは、採用分布形の理論クオンタイルと標本順

序統計量との誤差割合を測る指標で、次式によって定義される。

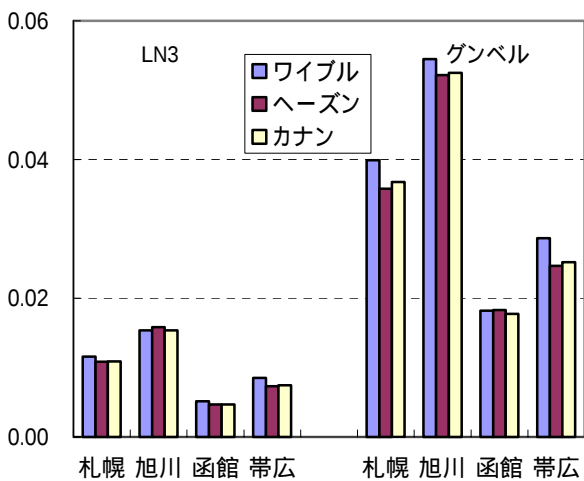
$$SLSC = \frac{\sqrt{\xi^2}}{|s_{0.99} - s_{0.01}|} \quad \xi^2 = \frac{1}{N} \sum_{i=1}^N (s_i - r_i)^2$$

ここに $s_{0.99}$ 、 $s_{0.01}$ ：それぞれの非超過確率 0.99 と 0.01 に対する当該確率分布の標準偏差、 s_i ：順序統計量データを推定母数で変換した標準変量、 r_i ：プロットングポジションに対応した理論クオントイルを推定母数で変換した標準変量。

まずプロットングポジションの公式の違いによる確率値の比較を行う。図—2 は、前途のワイブル、ヘーズン、カナンの 3 式について、道内 4 観測所の 1/150 確率値を LN3 およびグンベル分布で算出したものである。同じく図—3 はそれぞれの場合の SLSC を算出して比較している。



図—2 プロット法の違いによる確率値の比較

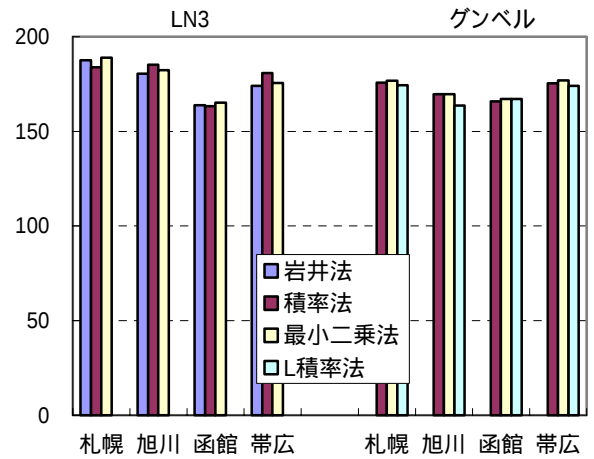


図—3 プロット法の違いによるSLSCの比較

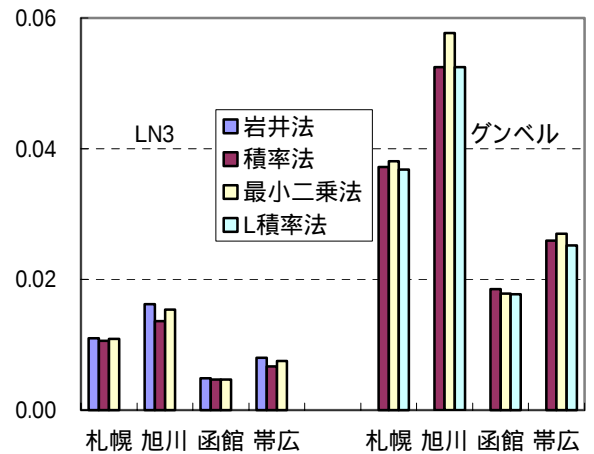
これらのグラフより、LN3・グンベルともにプロットングポジションの違いによる 150 年確率値及び SLSC の差は小さかった。そのため、ここからはプロットングポジションは、平均的な値を示すカナン式を代表して扱うことにする。

次に母数推定法の違いによる 150 年確率値の比較を図

—4 に、SLSC の比較を図—5 に示す。



図—4 母数推定法の違いによる確率値の比較



図—5 母数推定法の違いによるSLSCの比較

図—4、図—5 より、LN3・グンベル分布ともに多少ばらつきがあるものの、各地点で 150 年確率値及び SLSC の母数推定法による差は小さく、最小二乗法で精度は十分なことから、極値付近のデータのみを使った確率計算を最小二乗法で行うものとする。

4. 確率値付近のデータによる確率分布の適合度

前節までの検討結果から、ここではカナン式のプロットによる最小二乗法を用いて、降水量が大きいほうから 50%、30%のデータを使う場合の確率値の精度を検証する。最小二乗法で大きいほうから 50%のデータで母数を求める場合、各観測値のプロットングポジションは全データの場合と同じとする。これにより各データの確率的意味は変えないで母数を求められる。

50%のデータから母数を求めたものと全データから母数を求めたもので 50%データに対する適合度を比較すると、50%データから母数を求めたほうが適合度が向上すると考えられる。同様に 30%のデータから母数を求めたものは全データから母数を求めたものよりも 30%のデータに対する適合度が上昇すると思われる。

図—6、図—7 の確率値をみると、場所により差があ

るが、やや確率値が大きくなる傾向がみられ、第1位、第2位が大きな値の影響が強まっていると考えられる。

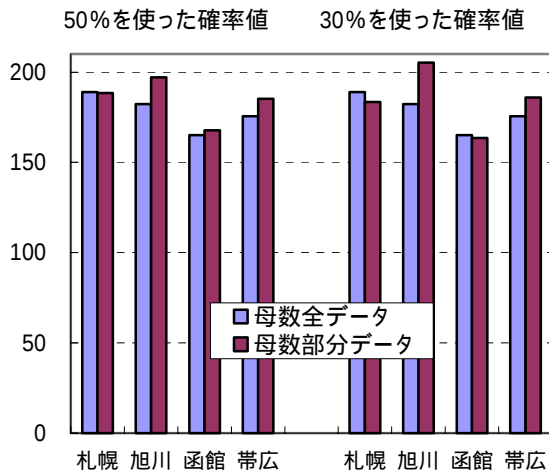


図-6 確率値の比較 (LN3)

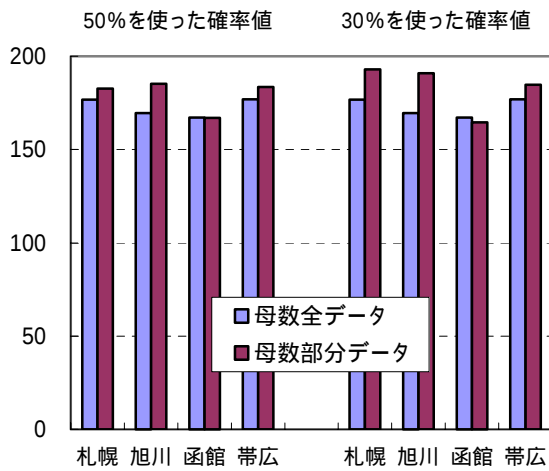


図-7 確率値の比較 (グンベル)

30%、50%部分の適合度を SLSC でみたものが図-8、図-9である。30%、50%の部分データで母数を求めたほうが、全データのものより小さく適合度が向上しているが、その割合はデータによる差が大きい。大きいほうから50%、30%のデータのSLSCは次式で表される。

$$50\% \text{の場合 } SLSC = \frac{\sqrt{\xi^2}}{|s_{0.99} - s_{0.50}|} \quad \xi^2 = \frac{1}{0.5N} \sum_{i=0.5N}^N (s_i - r_i)^2$$

$$30\% \text{の場合 } SLSC = \frac{\sqrt{\xi^2}}{|s_{0.99} - s_{0.70}|} \quad \xi^2 = \frac{1}{0.7N} \sum_{i=0.7N}^N (s_i - r_i)^2$$

50%を使ったSLSC 30%を使ったSLSC

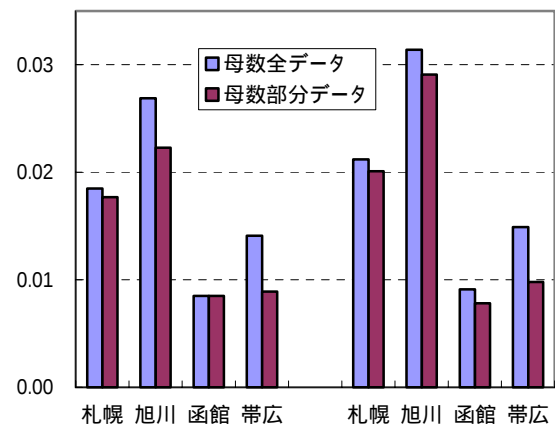


図-8 SLSCの比較 (LN3)

50%を使ったSLSC 30%を使ったSLSC

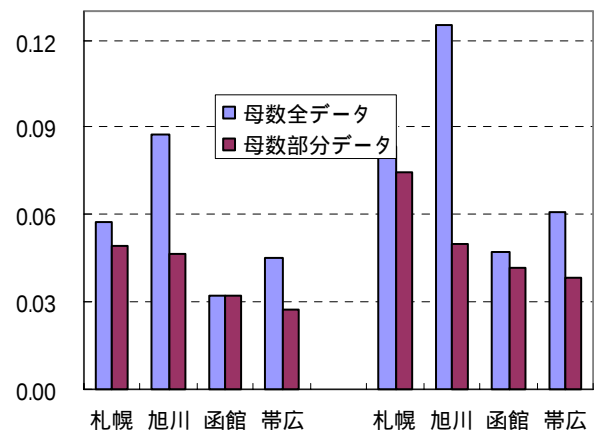


図-9 SLSCの比較 (グンベル)

同様に χ^2 分布による度数分布の棄却検定を行った場合の結果が、表-1である。ほとんどの場合、部分データで母数を算定した方が適合度が増加する。しかし例外的に札幌のLN3の30%と函館のグンベル30%では適合度が悪くなっている。

表-1 χ^2 検定 (危険率5%)

χ^2		LN3	グンベル	危険限界
札幌	母数全データ	6.27	10.61	16.90
	母数全データの50%	5.94	8.42	14.10
	母数部分データの50%	5.78	6.17	
	母数全データの30%	6.86	8.98	12.60
	母数部分データの30%	7.41	7.11	
旭川	母数全データ	5.07	9.18	15.51
	母数全データの50%	5.30	8.30	14.07
	母数部分データの50%	4.20	5.89	
	母数全データの30%	5.75	5.63	12.59
	母数部分データの30%	3.50	2.93	
函館	母数全データ	3.81	3.50	15.51
	母数全データの50%	3.80	3.72	12.59
	母数部分データの50%	3.37	3.63	
	母数全データの30%	7.84	6.98	11.07

	母数部分データの 30%	7.00	7.00	
帯 広	母数全データ	5.95	5.70	15.51
	母数全データの 50%	5.60	4.86	12.59
	母数部分データの 50%	3.81	3.56	
	母数全データの 30%	3.59	3.19	11.07
	母数部分データの 30%	2.41	2.36	

最後に対数正規分布を使い札幌のデータについて F 分布による棄却検定を行ったものを図-10 に示す。

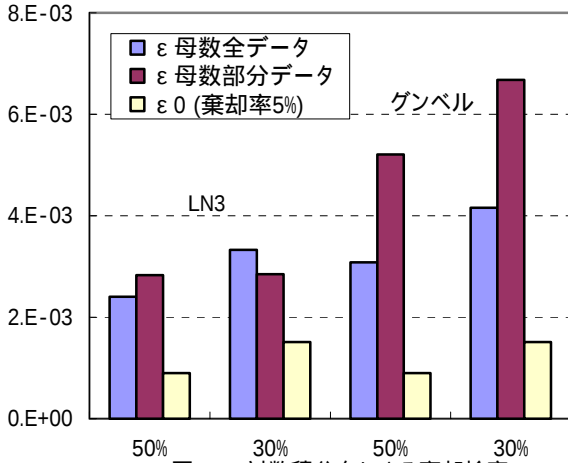


図-10 対数積分布による棄却検定

図に示すとおり、すべての場合で棄却できないが、全データの母数を使うより 50%または 30%の標本を使い母数を求めたほうが、棄却されにくい傾向になっている。

5. あとがき

確率分布の極値に近い部分の適合度を上げるための検討として、プロットングポジションを用いた最小二乗法により、極値に近い部分のデータによる確率値の計算法を検討した。

これを用いて大きいほうから 50%、30%のデータの範囲で SLSC や棄却検定を行い、50%、30%データで母数を算定した方が全データで母数を算定したものより、50%、30%の範囲の適合度が向上することを確認した。

参考文献

- 1) 星清：開発土木研究所月報別刷，北海道開発局開発土木研究所 1998.5 No540 P31～P78
- 2) 岩井重久，石黒政儀：応用水門統計学 P5～P38、P64～P136
- 3) 荒木正夫，椿東一郎：水理学演習下巻 P138～P146
- 4) 板元慶行，石黒真木夫，北川原四郎：情報量統計学 P27～P41

参考

岩井法 (LN3) 岩井法では、まず a を決める。

$$a' = \frac{x_{(1)} \cdot x_{(N)} - x_g^2}{x_{(1)} + x_{(N)} - 2x_g}$$

$$\text{観測値のメディアン } x_g : \log x_g = \frac{1}{N} \sum_{i=1}^N \log x_i$$

$x_{(1)}$ は標本の小さい値から、 $x_{(N)}$ は標本の大きい値から順に代入する。これを $m(\approx N/10)$ 回、計算をする。

$$a = \frac{1}{m} \sum_{i=1}^m a'$$

この a を次の式に代入し、母数を求める。

$$\mu_y = \frac{1}{N} \sum_{i=1}^N \ln(x_i - a) \quad \sigma_y^2 = \frac{1}{N} \sum_{i=1}^N [\ln(x_i - a) - \mu_y]^2$$

積率法 (LN3) 次の式を使い母数を求める。

$$\bar{x} = \mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad S^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$M_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3 \quad Cs = \frac{M_3}{S^3}$$

$$\sigma_x = \left(\frac{N}{N-1} \right)^{\frac{1}{2}} S \quad \gamma_x = Cs(A + B \cdot Cs^3)$$

$$A = 1.01 + 7.01/N + 14.66/N^2 \quad B = 1.69/N + 74.66/N^2$$

$$\beta = 1 + \left(\frac{\lambda_x^2}{2} \right) \quad \phi = \left[\beta + (\beta^2 - 1)^{\frac{1}{2}} \right]^{\frac{1}{3}} + \left[\beta - (\beta^2 - 1)^{\frac{1}{2}} \right]^{\frac{1}{3}} - 1$$

$$\sigma_y = (\ln \phi)^{\frac{1}{2}} \quad \varepsilon = \frac{\sigma_x}{[\phi(\phi - 1)]^{\frac{1}{2}}}$$

$$\mu_x = \ln \varepsilon \quad a = \mu_x - \varepsilon \cdot \phi^{\frac{1}{2}}$$

積率法(グンベル) 次の式を使い母数を求める。

$$\mu_x = c + 0.5772a \quad \sigma_x = \frac{\pi a}{\sqrt{6}} \quad \gamma_x = 1.1396$$

L 積率法(グンベル) 確立重みつき積率の標本推定値を計算する。

$$\beta_0 = \frac{1}{N} \sum_{i=1}^N x_i \quad \beta_1 = \frac{1}{N(N-1)} \sum_{i=1}^N (i-1)x_i$$

$$\beta_1 = \frac{1}{N(N-1)(N-2)} \sum_{i=1}^N (i-1)(i-2)x_i$$

これらを λ に換える。

$$\lambda_1 = \beta_0 \quad \lambda_2 = 2\beta_1 - \beta_0 \quad \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$$

ここで L 積率と母数の関係は次式なので、以下より母数を求める。

$$\lambda_1 = c + 0.5772a \quad \lambda_2 = a \ln 2$$