II-74　極値分布の母数推定法の比較評価

京都大学工学部　正員　宝　　馨
京都大学工学部　正員　高棹琢馬
大　　阪　　府　正員　清水　章

1．緒言　Gumbel分布，一般化極値(Generalized Extreme Value, GEV)分布は，極値理論に基づく分布として極値データの頻度解析に有用である。これらそれぞれについて提案されている種々の母数推定法を，極値水文量を想定したモンテカルロ・シミュレーションにより，再現確率統計量（確率水文量）の平均二乗誤差（Mean Square Error, MSE）を評価規準として，比較検討する。

2．累積分布関数　Gumbel分布:$F(x) = \exp\{-\exp(s)\}$; $s = a(x - b)$:標準変量 a，b:尺度母数，位置母数。GEV分布:$F(x) = \exp[-\{1 - k\cdot(x - \xi)/\alpha\}^{1/k}]$; α，ξ，k:それぞれ尺度母数，位置母数，形状母数。GEV分布は，k=0のときGumbel分布と，k>0，k<0のときそれぞれ対数極値分布A型，B型と等しい。

3．母数推定法　①最尤法→最適化プログラムを用いて尤度関数を最大化；　②最小二乗法→標準変量に対して最小二乗法を適用，プロッティング・ポジション公式としてHazen公式，Weibull公式の２つを検討；③積率法→標本積率を理論積率と等しく置いて母数を推定，GEV分布は歪係数として標本歪係数，不偏歪係数の２つを検討；　④PWM(= Probability Weighted Moments)法[1]　→非超過確率の累乗で加重した積率であるPWMを用いて母数を推定；　⑤最大エントロピー法[2]　→エントロピーを最大とする確率分布モデルについて，その制約条件から母数を推定；　Gumbel分布は①～⑤，GEV分布は①,③,④の母数推定法を検討する。

4．平均二乗誤差　ある統計量(母数，確率水文量など)をθ，その推定値を$\hat{\theta}$とするとき，平均二乗誤差は次式で定義される。　$MSE = E[(\hat{\theta} - \theta)^2] = (E[\hat{\theta}] - \theta)^2 + E[(\hat{\theta} - E[\hat{\theta}])^2] = BIAS^2 + SD^2$
すなわち，MSEは推定値の偏倚(BIAS)と推定誤差(SD)を併せて評価することができる。
以下では，MSEの平方根をとったものをRMSE(root mean square error)と記し，これを評価規準とする。

5．方法　①　確率分布モデルの母数を想定する。　②　０～１の一様乱数と累積分布関数の逆関数を用いて，想定した確率分布に従う乱数をN個発生させる。　③　②の標本について母数推定値と確率水文量を算定する。④　②～③の作業をM回繰り返し確率水文量のRMSEを求めて，RMSEの小さい母数推定法を良いと結論する。また，⑤　Nの値を変えて②～④の作業を行うと，標本数の増加に伴うRMSEの推移がわかる。

今回のシミュレーションは，母分布として大阪の年最大日降水量(1889年～1980年)と St. Marys川の年最大日流量(1916年～1974年)に近い２つの場合を用いたが，どちらもほぼ同じ評価を示した。以下では，大阪の場合についてのみ述べることにする。手順①の母数は，最尤法を用いて推定した母数に近い値（Gumbel分布;a=0.04，b=77.0，GEV分布;α=20.0，ξ=75.0，k=-0.1）を与えた。また，手順②の標本数Nは 10，20，･･･，100，200，500，1000，手順④の繰り返し回数 M=5000 とした。

6．結果と考察　表－１，表－２は，Gumbel分布と GEV分布の100年確率水文量の BIASとRMSEを示したものである。これらの表から以下のことが言える。

【Gumbel分布の場合】　RMSEはNの値にかかわらず最尤法，最大エントロピー法，PWM法の順に小さい。BIASは N=1000のときRMSEと同じ順であるが，30≦N≦500のときPWM法，積率法，最大エントロピー法の順に小さい。特に，PWM法のBIASは極めて小さい。また，積率法はBIAS，RMSEのどちらも最小二乗法より小さい。

【GEV分布の場合】　RMSEはN=1000のとき最尤法，PWM法の順に小さいが，N<1000のときPWM法，最尤法の順に小さい。BIASはRMSEと同じ順位であるが，PWM法のBIASはN=70程度で1mm未満と極めて小さい。また N≦20のとき積率法のRMSEは最尤法より小さい。標本歪係数C_sはalgebraic boundedness[3]と呼ばれる上下限がある。

$$-(N - 2)/(N - 1)^{1/2} < C_s < (N - 2)/(N - 1)^{1/2}, \quad C_s = \Sigma_i(x_i - \bar{x})^3/N \;/\; \{\Sigma_i(x_i - \bar{x})^2/N\}^{2/3}$$

このため，Nが小さいとき積率法の推定誤差は小さくなると思われる。

7．結論 (1) MSEでみると Gumbel分布は最尤法が，GEV分布はPWM法が良い。 (2) どちらの分布についても PWM法のBIASは極めて小さく，データ数が約70個以上あれば，190mm程度の100年確率水文量に対してBIASは 1mm未満である。 (3) MSEでみると Gumbel分布の場合，積率法は最小二乗法に優る。 また，GEV分布の場合，小標本（N=10,20）のとき 積率法は最尤法に優る。 歪係数の algebraic boundedness により推定値の変動が抑えられるためであると思われる。

表-1 Gumbel分布に対するデータ数Nと 100年確率水文量の平均二乗誤差の関係
（上段：推定値，下段：平均二乗誤差の平方根，M＝5000，真値=192.00）

母数推定法	N = 10	N = 20	N = 30	N = 50	N = 70	N = 100	N = 200	N = 500	N =1000
歪：平均値	0.6275	0.7646	0.8366	0.9244	0.9695	1.0156	1.0748	1.1104	1.1250
標準偏差	(0.4328)	(0.5028)	(0.5064)	(0.4944)	(0.4555)	(0.4194)	(0.3341)	(0.2331)	(0.1714)
最尤法	184.31	187.59	189.13	190.31	190.85	191.88	191.50	191.76	191.90
	(33.27)	(23.49)	(18.72)	(14.41)	(12.14)	(9.94)	(7.11)	(4.49)	(3.18)
最小二乗法	282.31	197.20	195.66	194.52	193.87	193.33	192.83	192.37	192.23
(Hazen)	(43.81)	(30.02)	(23.94)	(18.49)	(15.51)	(12.61)	(9.01)	(5.68)	(4.02)
最小二乗法	228.93	213.57	208.04	203.23	200.76	198.71	196.11	194.05	193.22
(Weibull)	(63.72)	(40.23)	(30.95)	(22.90)	(18.69)	(14.88)	(10.19)	(6.14)	(4.25)
積率法	192.53	191.00	190.97	191.20	191.30	191.37	191.68	191.82	191.92
	(38.39)	(27.48)	(22.36)	(17.51)	(14.88)	(12.17)	(8.80)	(5.60)	(3.99)
PWM 法	196.64	193.15	192.49	192.15	192.03	191.87	191.92	191.93	191.97
	(37.77)	(26.06)	(20.83)	(16.03)	(13.60)	(11.00)	(7.88)	(4.99)	(3.57)
最大ｴﾝﾄﾛﾋﾟｰ	187.11	188.87	189.87	190.70	191.06	191.22	191.58	191.80	191.91
法	(34.09)	(24.09)	(19.23)	(14.80)	(12.52)	(10.10)	(7.29)	(4.61)	(3.28)

表-2 GEV 分布に対するデータ数Nと 100年確率水文量の平均二乗誤差の関係
（上段：推定値，下段：平均二乗誤差の平方根，M＝5000，真値=191.82）

母数推定法	N = 10	N = 20	N = 30	N = 50	N = 70	N = 100	N = 200	N = 500	N =1000
歪：平均値	0.7418	0.9936	1.1453	1.3347	1.4363	1.5362	1.6819	1.7955	1.8477
標準偏差	(0.4882)	(0.6194)	(0.6590)	(0.7055)	(0.6982)	(0.6920)	(0.6531)	(0.5770)	(0.4854)
最尤法	238.88	230.54	213.43	209.04	201.17	199.26	192.45	192.03	191.81
	(159.04)	(127.57)	(100.73)	(85.93)	(59.77)	(52.54)	(19.70)	(12.22)	(7.10)
積率法	93.17	107.35	118.10	132.18	141.88	151.87	170.60	186.07	190.11
(標本歪)	(130.08)	(118.40)	(109.55)	(97.75)	(88.66)	(78.34)	(56.04)	(26.50)	(11.46)
積率法	108.75	114.93	123.21	135.73	144.54	154.08	171.40	186.25	190.14
(不偏歪)	(121.63)	(113.66)	(106.07)	(95.07)	(86.43)	(76.23)	(55.02)	(26.12)	(11.45)
PWM 法	208.84	197.32	194.71	193.36	192.68	192.21	192.04	191.82	191.84
	(76.55)	(53.36)	(43.18)	(33.85)	(28.24)	(23.03)	(16.39)	(10.26)	(7.30)

1) J.A. Greenwood et al., Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form, W.R.R., 15(5), 1974, pp.1049-1054.
2) P.W. Jowitt, The Extreme Value type 1 Distribution and the Principle of Maximum Entropy, J. Hydrology, 42, 1979, pp.23-38.
3) K. Kirby, Algebraic Boundedness of Sample Statistics, W.R.R., 10(2), 1974, pp.220-222.